بسم الله الرحمن الرحیم

کارگاه علم داده: مدل سازی داده (۳)

Topology and Geometry of Stochastic field

سیدمحمدصادق موحد

دانشکده فیزیک - دانشگاه شهید بهشتی

پژوهشکده فیزیک، پژوهشگاه دانشهای بنیادی (IPM)

گروه کیهان شناسی محاسباتی (GCC-SBU)

آزمایشگاه میان رشته ای ابن سینا

http://facultymembers.sbu.ac.ir/movahed/

۹۸/۶/۷

تبریز

IPM
Institute for Research in Fundamental Sciences

دانشگاه شهید بهشتی
دانشکده فیزیک

فهرست مطالب

مهمترین منابع:

- Algebraic Topology of Random Fields and Complexes, Research thesis, Omer Bobrowski, 2012
- Random Fields and Geometry, Adler, R. J., Taylor, Jonathan E., Springer 2007.
- The Geometry of Random fields, Adler, 1981.
- GEOMETRY, TOPOLOGY and Physics, M. Nakahara, 2003
- Topological Complexity of Smooth Random Functions, Robert J. Adler, Jonathan E. Taylor, Springer 2009.
- Matsubara, Takahiko. "Statistics of smoothed cosmic fields in perturbation theory. I. Formulation and useful formulae in second-order perturbation theory." The Astrophysical Journal 584.1 (2003): 1.
- Ghasemi Nezhadhaghighi, M., et al. "Characterization of the anisotropy of rough surfaces: Crossing statistics." Journal of Applied Physics 122.8 (2017): 085302.
- Pirlar, M. Arshadi, et al. "Crossing statistics of laser light scattered through a nanofluid." JOSA A 34.9 (2017): 1620-1631.
- **http://facultymembers.sbu.ac.ir/movahed/**

بخش اول: داده هایی با ماهیت تصادفی

Part 1
Stochastic fields
Stochastic processes
Random fields

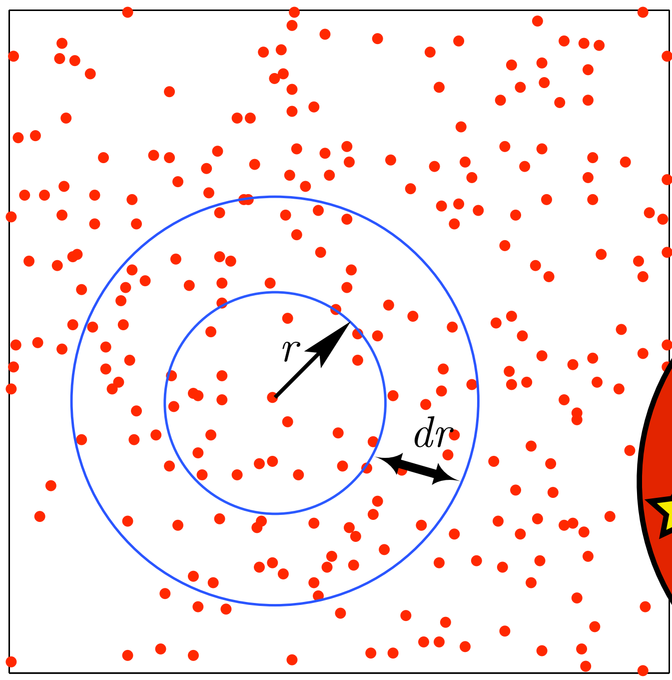
$$\{\mathcal{F}, T\} \in \mathcal{M}$$

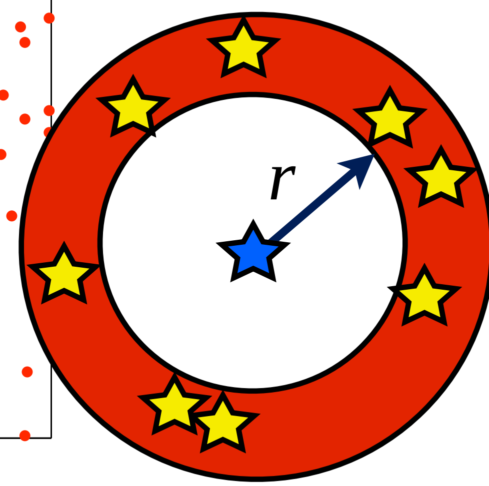
$$\mathcal{F}^d : \Omega \to \mathbb{R}^T$$

$$T \subset \mathbb{R}^N$$

$\mathcal{F}$ is a (N, d)-stochastic(random) field

Why stochastic field?





Initial conditions and evolution

① Mass-Conservation or $\mathcal{L} f = C[f]$

$$\frac{D\rho}{Dt} = -\rho \vec{\nabla}\cdot\vec{v}$$

② Momentum Conservation

$$\dot{\vec{v}} + (\vec{v}\cdot\vec{\nabla})\vec{v} = -\vec{\nabla}\phi - \frac{\vec{\nabla}P}{\rho}$$

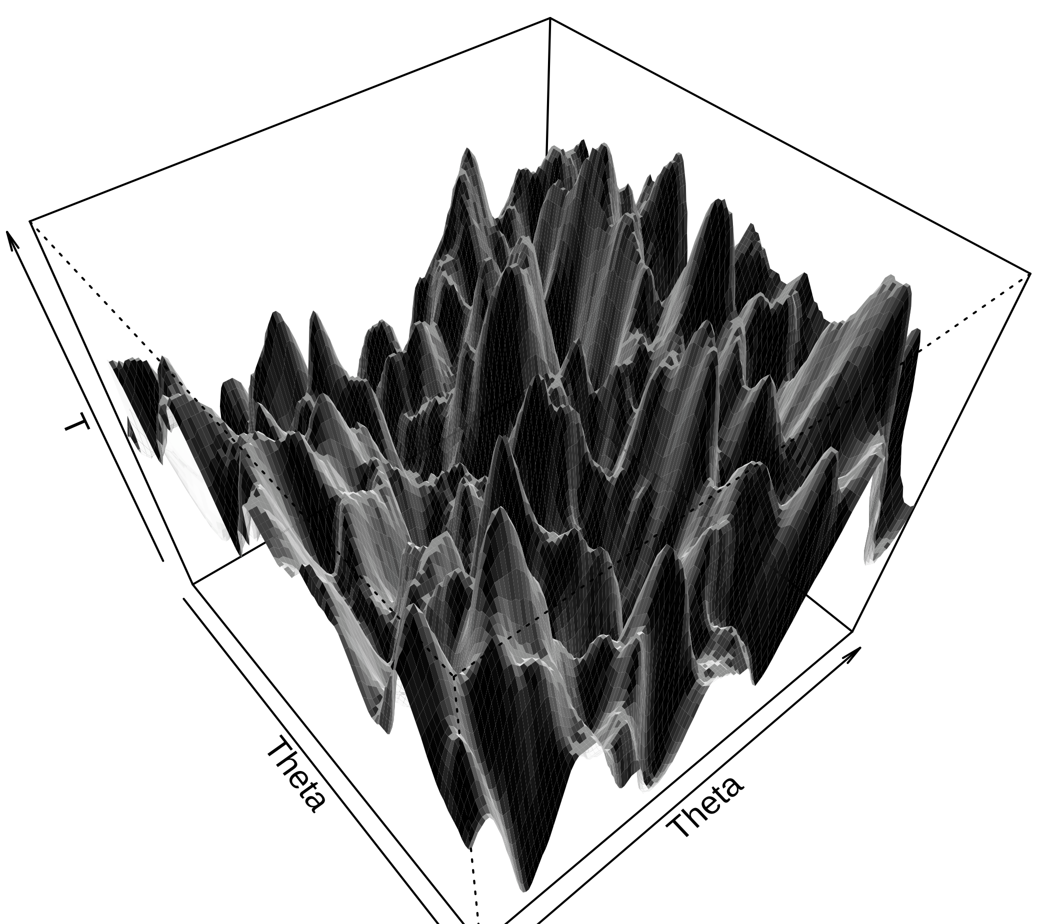
③ Gravitational instability : Poisson Eq.

$$\nabla^2\phi = 4\pi G\rho$$

④ Entropy Conservation

$$\frac{DS}{Dt} = \frac{\partial S}{\partial t} + (\vec{v}\cdot\vec{\nabla})S = 0$$

⑤ Equation of State $P \to \rho$



Evolution equation of a stochastic field

Suppose that $\mathcal{F}$ is a (N, d) stochastic field. The independent parameter is N-dimensional parameter called T.

$$\frac{\partial P(\{\mathcal{F}\};\{T\})}{\partial T_j} = \mathcal{L}_j P(\{\mathcal{F}\};\{T\})$$

$$\mathcal{L}_j \equiv \sum_{\nu=1}^{\infty} \frac{(-\partial)^\nu}{\partial \mathcal{F}_{i_1} \partial \mathcal{F}_{i_2} \ldots \partial \mathcal{F}_{i_\nu}} D^{(\nu)}_{i_1, i_2 \ldots i_\nu}(\{\mathcal{F}\}; T_j)$$

If $D^{(\nu)} = 0$ for $\nu \geq 3$ therefore

$$\frac{\partial \mathcal{F}_i}{\partial T_j} = D^{(1)}_i(\{\mathcal{F}\}; T_j) + \sqrt{D^{(2)}_{ik}(\{\mathcal{F}\}; T_j)}\,\eta_k(\{\mathcal{F}\}; T_j)$$

where

$$\langle \eta_i(\{\mathcal{F}\};\{T\}) \eta_j(\{\mathcal{F}\};\{T\}) \rangle = \delta_{ij} \delta_D(\mathcal{F}_i, \mathcal{F}_j)$$

Ornstein-Uhlenbeck Equation

$$\dot{x}_i(t) + \sum_{j=1}^{M} \xi_{ij}\, \dot{x}_j = \eta_i(t) \qquad \langle \eta_i(t)\, \eta_j(t') \rangle = \delta_{ij}\, \delta_D(t-t')$$

$$\dot{v}(t) = -\xi\, v(t) + \eta(t)$$ → Stationary regime → M.B. Distribution

$$\dot{x}(t) = \int G(t,t')\, \xi(t';x)\, dt' + K(x,t)\, \eta(t)$$

$$\langle \eta(t)\, \eta(t') \rangle = \delta_D(t-t')$$

Observables

1) Quantitative measures

2) Geometrical and Topological measures

3) Dual Space measures

$$\delta(t;\bar{X}) \equiv \frac{\rho(t;\bar{X}) - \langle \rho(t;\bar{X}) \rangle}{\langle \rho(t;\bar{X}) \rangle} \qquad \delta\Phi(t;\bar{X}) \equiv \frac{\Phi(t;\bar{X}) - \langle \Phi(t;\bar{X}) \rangle}{\langle \Phi(t;\bar{X}) \rangle}$$

$$\delta\vec{V}(t;\bar{X}) \equiv \frac{\vec{V}(t;\bar{X}) - \langle \vec{V}(t;\bar{X}) \rangle}{\langle \vec{V}(t;\bar{X}) \rangle} \qquad \delta T(t;\bar{X}) \equiv \frac{T(t;\bar{X}) - \langle T(t;\bar{X}) \rangle}{\langle T(t;\bar{X}) \rangle}$$

$$A_{\mu\nu\eta\ldots} = \Big(\alpha(r_\mu), \alpha(r_\mu)_{;1}, \alpha(r_\mu)_{;2}, \alpha(r_\mu)_{;3}, \alpha(r_\mu)_{;11}, \alpha(r_\mu)_{;22}, \alpha(r_\mu)_{;33}, \alpha(r_\mu)_{;12}, \alpha(r_\mu)_{;13}, \alpha(r_\mu)_{;23},$$
$$\Big(\alpha(r_\nu), \alpha(r_\nu)_{;1}, \alpha(r_\nu)_{;2}, \alpha(r_\nu)_{;3}, \alpha(r_\nu)_{;11}, \alpha(r_\nu)_{;22}, \alpha(r_\nu)_{;33}, \alpha(r_\nu)_{;12}, \alpha(r_\nu)_{;13}, \alpha(r_\nu)_{;23}, \ldots\Big)$$

$$\langle \mathcal{F}(A) \rangle = \int d^N A \mathcal{F}(A) P(A)$$

Probability density function of features in an arbitrary smoothed stochastic field
Data is considered as regular sampled

Perturbative expansion of Statistics 1

$$f \rightarrow f' \equiv f - \langle f \rangle \quad \rightarrow \langle f' \rangle = 0 \qquad \sigma_0^2 = \langle f^2 \rangle = \frac{1}{(2\pi)^{d/2}} \int d^d k P(k) \quad \alpha \equiv \frac{f}{\sigma_0}$$

$$A_{\mu\nu\eta\ldots} = \big(\alpha(r_\mu), \alpha(r_\mu)_{;1}, \alpha(r_\mu)_{;2}, \alpha(r_\mu)_{;3}, \alpha(r_\mu)_{;11}, \alpha(r_\mu)_{;22}, \alpha(r_\mu)_{;33}, \alpha(r_\mu)_{;12}, \alpha(r_\mu)_{;13}, \alpha(r_\mu)_{;23},$$
$$\big(\alpha(r_\nu), \alpha(r_\nu)_{;1}, \alpha(r_\nu)_{;2}, \alpha(r_\nu)_{;3}, \alpha(r_\nu)_{;11}, \alpha(r_\nu)_{;22}, \alpha(r_\nu)_{;33}, \alpha(r_\nu)_{;12}, \alpha(r_\nu)_{;13}, \alpha(r_\nu)_{;23}, \ldots\big)$$

$$Z_A(\lambda) \equiv \langle \exp(i\lambda \bullet A) \rangle_A = \int_{-\infty}^{+\infty} d^N A\, P(A) \exp(i\lambda \bullet A)$$

Moment

$$= 1 + \sum_{n=1} \frac{i^n}{n!} \left(\sum_{\mu_1=1}^{N} \sum_{\mu_2=1}^{N} \cdots \sum_{\mu_a=1}^{N} \sum_{\nu_1=1}^{N} \sum_{\nu_2=1}^{N} \cdots \sum_{\nu_b=1}^{N} M^{(a+b=n)}_{\mu_1\mu_2..\mu_a;\nu_1,\nu_2..\nu_b} \lambda_{\mu_1} \lambda_{\mu_2} \ldots \lambda_{\mu_a} \lambda_{\nu_1} \lambda_{\nu_2} \ldots \lambda_{\nu_b} \right)$$

$$\ln(Z_A(\lambda)) = \sum_{n=1} \frac{i^n}{n!} \left(\sum_{\mu_1=1}^{N} \sum_{\mu_2=1}^{N} \cdots \sum_{\mu_a=1}^{N} \sum_{\nu_1=1}^{N} \sum_{\nu_2=1}^{N} \cdots \sum_{\nu_b=1}^{N} K^{(a+b=n)}_{\mu_1\mu_2..\mu_a;\nu_1,\nu_2..\nu_b} \lambda_{\mu_1} \lambda_{\mu_2} \ldots \lambda_{\mu_a} \lambda_{\nu_1} \lambda_{\nu_2} \ldots \lambda_{\nu_b} \right)$$

Cumulant

Free energy

Perturbative expansion of Statistics 2

$$\langle F \rangle_A = \int_{-\infty}^{+\infty} d^N A \, P(A) F$$

$$\langle F \rangle_A = \left\langle \exp\left(\sum_{n=3}^{\infty} \frac{(-1)^n}{n!} \left(\sum_{\mu_1=1}^{N} \sum_{\mu_2=1}^{N} \cdots \sum_{\mu_a=1}^{N} \sum_{\nu_1=1}^{N} \sum_{\nu_2=1}^{N} \cdots \sum_{\nu_b=1}^{N} K^{(a+b=n)}_{\mu_1\mu_2..\mu_a;\nu_1,\nu_2..\nu_b} \frac{\partial^n}{\partial A_{\mu_1} \ldots \partial A_{\mu_a} \partial A_{\nu_1} \ldots \partial A_{\nu_b}} \right) \right) F \right\rangle_G$$

$$\langle F \rangle_A = \langle F \rangle_G + \frac{1}{3!} \sum_{\mu_1=1}^{N} \sum_{\mu_2=1}^{N} \sum_{\mu_3=1}^{N} K^{(3)}_{\mu_1\mu_2\mu_3} \left\langle F_{;\mu_1\mu_2\mu_3} \right\rangle_G + \ldots.$$

$$F \equiv \delta(\alpha - \beta) \quad \alpha \equiv \frac{f}{\sigma_0}$$

$$P(\alpha) = \int dA_{\mu_2} dA_{\mu_3} \ldots dA_{\mu_N} \delta(\alpha - \beta) P(\vec{A})$$

$$P(f) = \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left(-\frac{\alpha^2}{2} \right) + \frac{1}{3!} K^{(3)}_{111} \left\langle \frac{\partial^3 \delta(\alpha - \beta)}{\partial \beta^3} \right\rangle_G + \ldots$$

Skewness

$$P(f) = \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left(-\frac{\alpha^2}{2} \right) \left[1 + \frac{1}{6} K^{(3)}_{111} H_3(\alpha) + \mathrm{O}(\sigma_0^3) \right]$$

Hermit polynomial

Perturbative expansion of Statistics III

$$Z_A(\lambda) \equiv \left\langle \exp\left(i\lambda \bullet A\right)\right\rangle_A = \int_{-\infty}^{+\infty} d^N A \, P(A) \exp\left(i\lambda \bullet A\right)$$

$$P(\vec{A}) = \frac{1}{(2\pi)^N} \int_{-\infty}^{+\infty} d^N \lambda \, Z_A(\lambda) \exp\left(-i\lambda \bullet A\right)$$

$$= \exp\left(\sum_{n=3}^{\infty} \frac{(-1)^n}{n!} \left(\sum_{\mu_1=1}^{N} \sum_{\mu_2=1}^{N} \cdots \sum_{\mu_a=1}^{N} \sum_{\nu_1=1}^{N} \sum_{\nu_2=1}^{N} \cdots \sum_{\nu_b=1}^{N} K^{(a+b=n)}_{\mu_1 \mu_2 .. \mu_a ; \nu_1 , \nu_2 .. \nu_b} \frac{\partial^n}{\partial A_{\mu_1} \ldots \partial A_{\mu_a} \partial A_{\nu_1} \ldots \partial A_{\nu_b}} \right) \right)$$

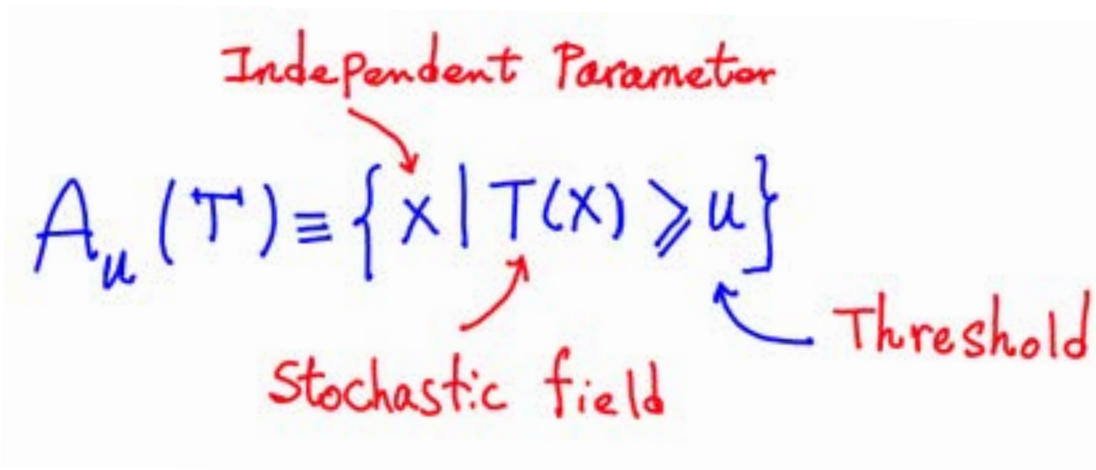
$$\times P_G(\vec{A})$$

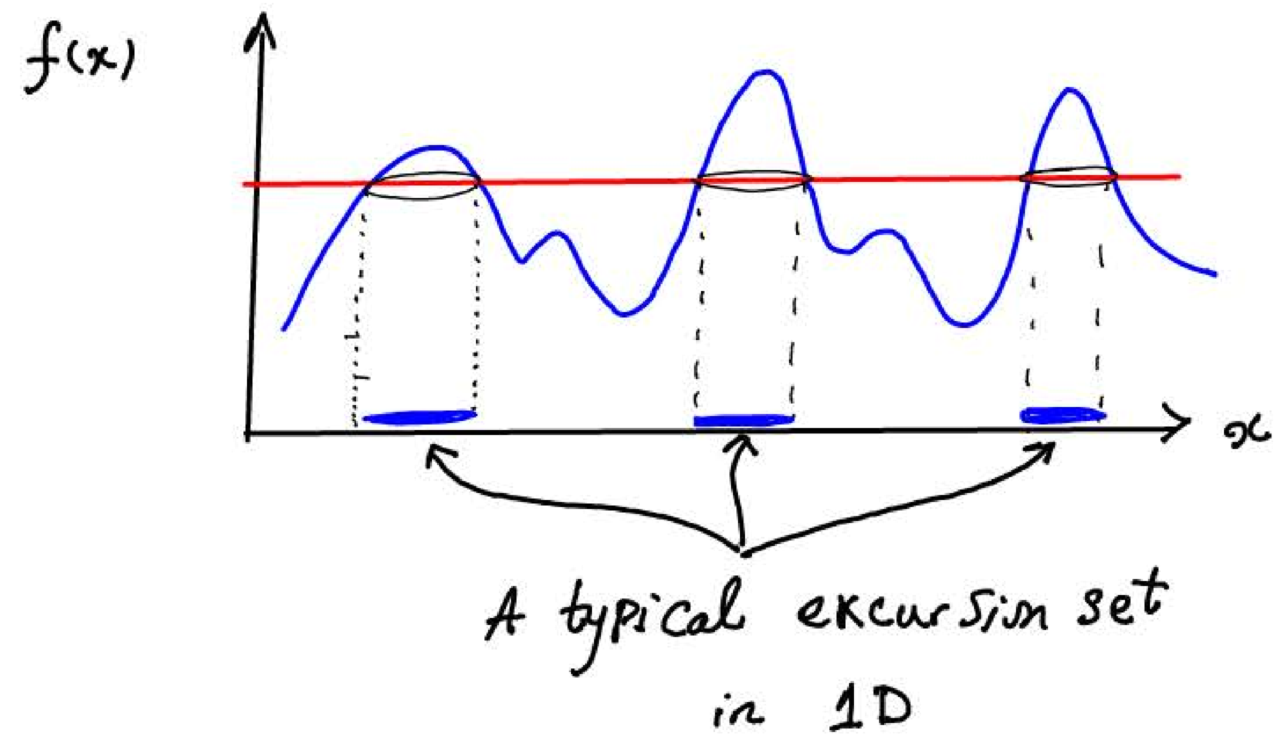
$$P_G(\vec{A}) = \frac{\exp\left(-\frac{1}{2} \vec{A}^T \bullet \left(K^{(2)}\right)^{-1} \bullet \vec{A} \right)}{(2\pi)^{N/2} \sqrt{Det\left|K^{(2)}\right|}}$$

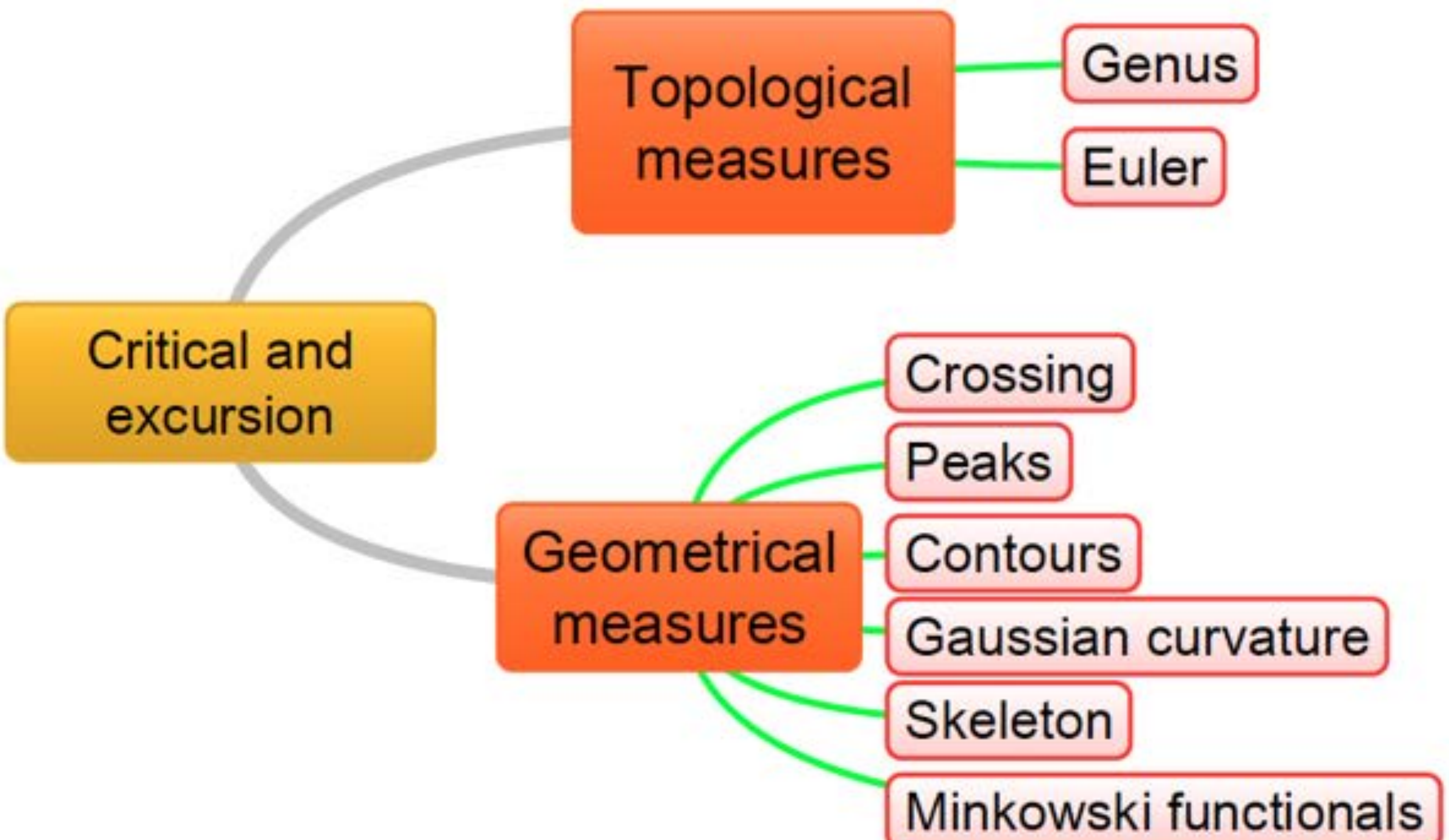
Covariance matrix or The inverse of Fisher information matrix

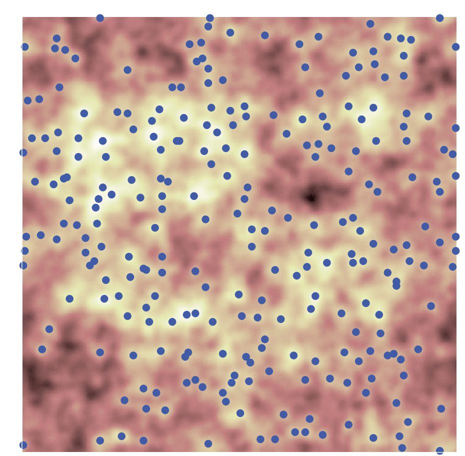
بخش دوم: خواص هندسی و توپولوژیک

General features and some proposed methods

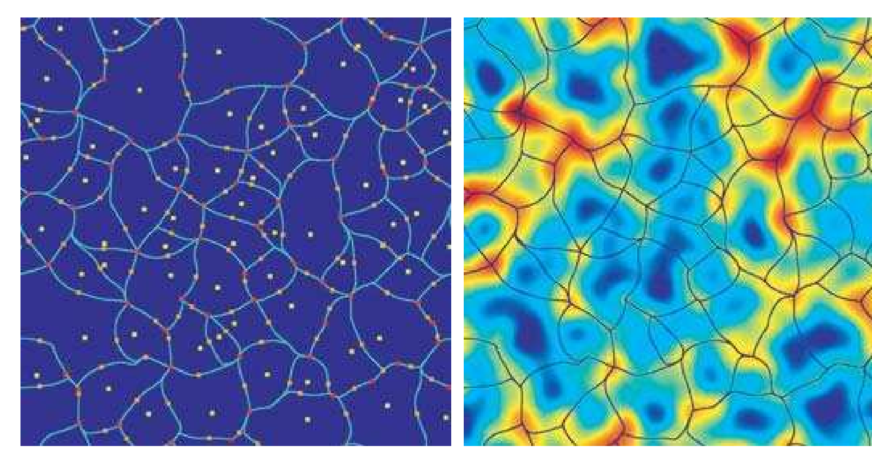








S.M.S.M., B. Javanmardi, R. K. Sheth, MNRAS, (2013)



Mon. Not. R. Astron. Soc. **366,** 1201–1216 (2006)

A brief about Topology

Topology is (roughly) the study of properties invariant under "continuous transformation

- Two shapes are topologically equivalent if and only if one shape can continuously deform to the other shape.
e.g. Sphere, cube, pyramid are all topologically equivalent.
On the other hands, Sphere and torus are different from topological point of view.

Why is topology so important?

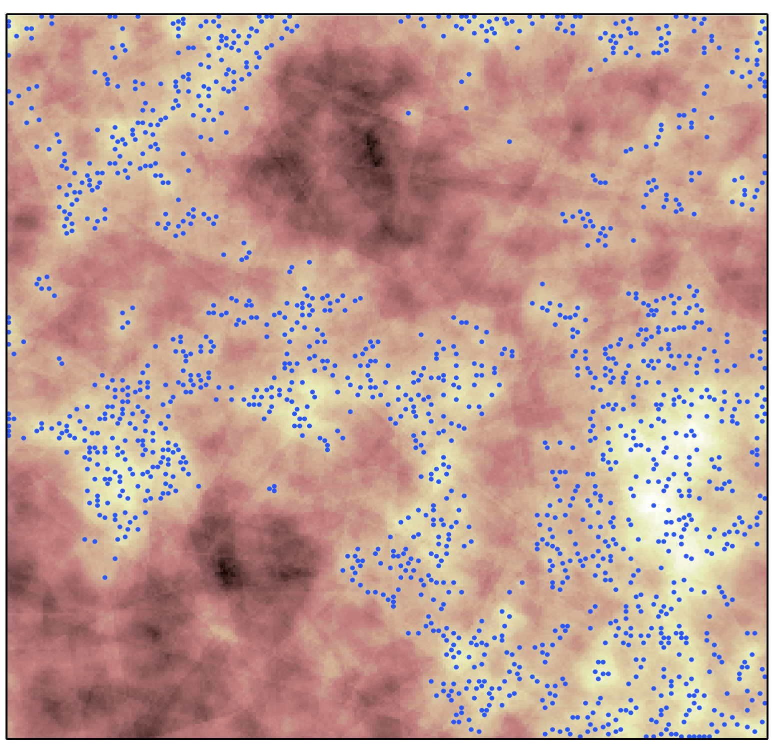
To answer to this question let me explain PDF and correlation function
- PDF shows the abundance of features
while
-correlation corresponds to probability of finding features with a condition

To distinguish between various stochastic fields mentioned tools are not enough

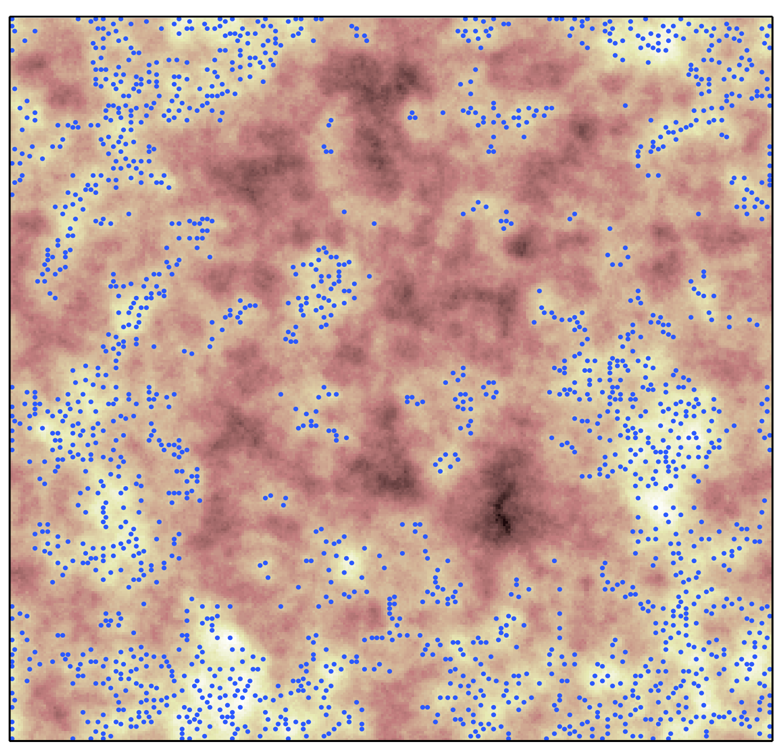
Why topological and geometrical measures?

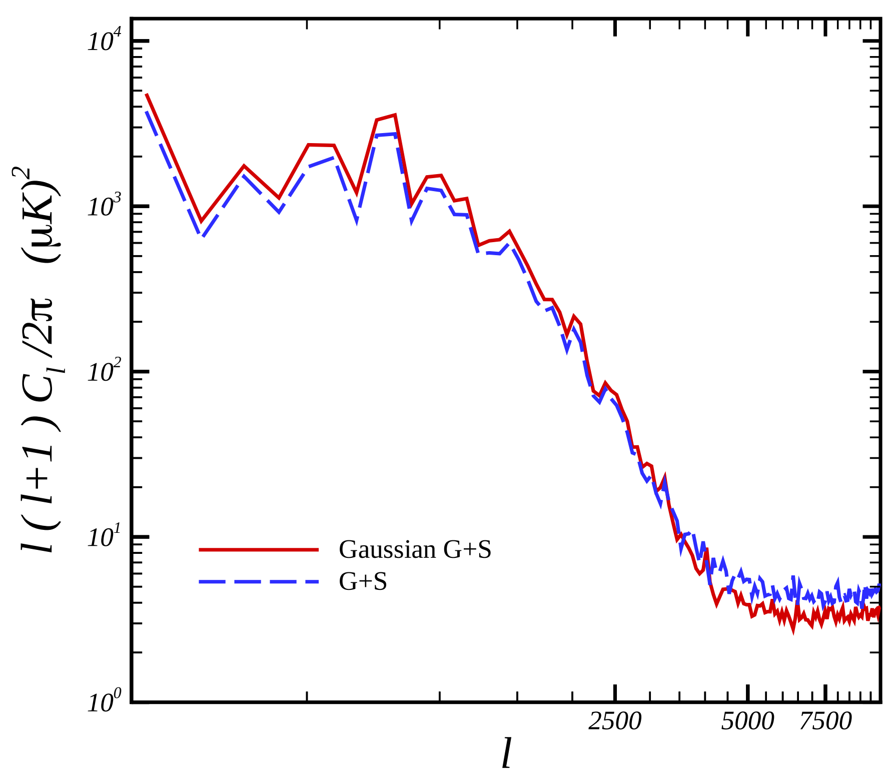
Both of these fields have same power spectrum
But their textures are completely different

GS



Gaussian-GS





Movahed et.al, MNRAS 2013

بخش سوم: مثال
Crossing statistics

Theoretical approach

One-point statistics

$$\langle f \rangle = \langle \text{Conditions correspond to feature} \rangle$$

$$= \langle f \rangle_{Gaussian} + \text{Perturbative Parts}\Big|_{NG + Anisotropy}$$

Two-point statistics

$$\langle f(r_1)\, g(r_2) \rangle = \int dA_1\, dA_2\, P(A_1, A_2)\, f(r_1)\, g(r_2)$$

$$P(\vec{A}_1, \vec{A}_2) = \left[\frac{1}{2\pi^N \mathrm{Det}(K)}\right]^{1/2} \exp\left(-\frac{A_1^{\dagger} \cdot K^{-1} \cdot A_2}{2}\right)$$

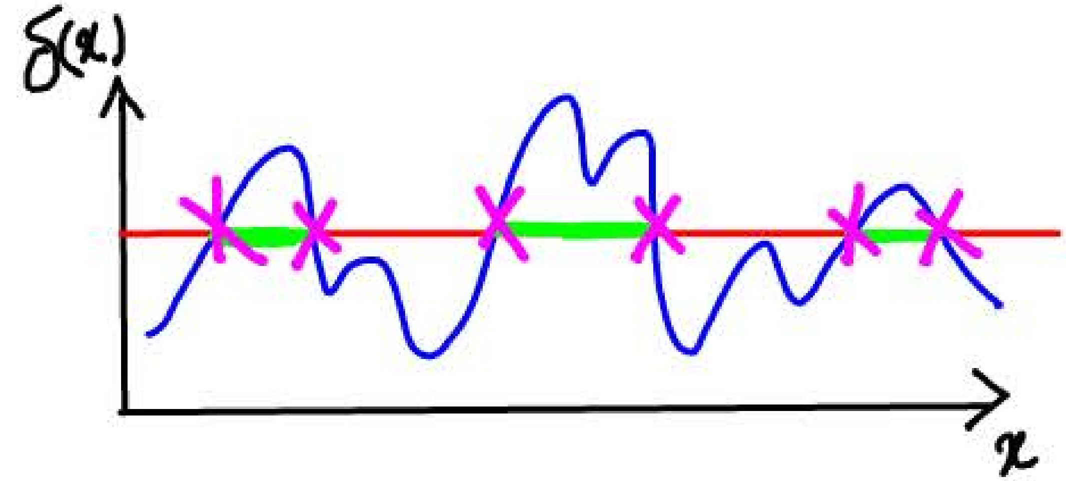
T. Matsubara, APJ 2003; S. Codis et. al., 1305.7402, Christophe Gay et. al., PRD 2012

Minkowski Functionals

1D field

$$V_0(\nu) = \int_Q d\ell = \langle \Theta(\alpha-\nu) \rangle$$

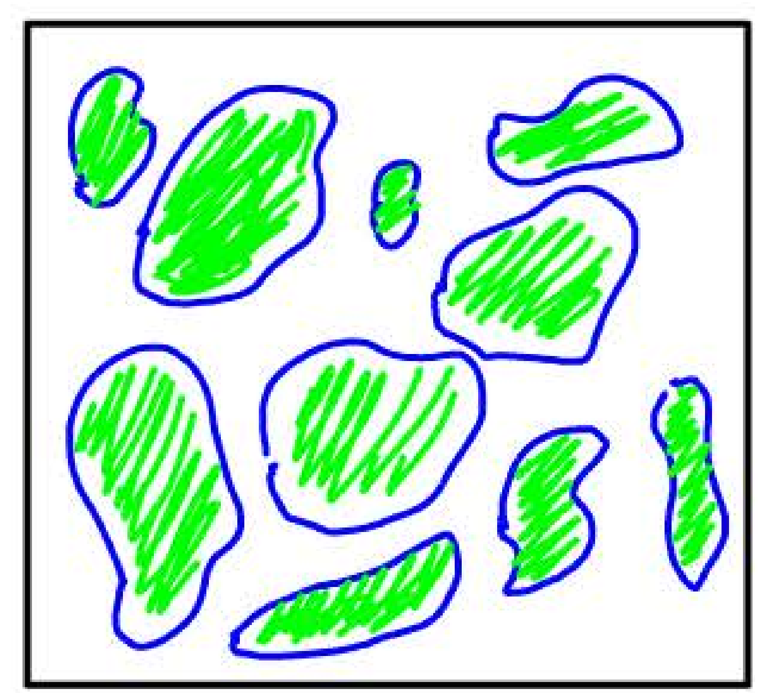
$$V_1(\nu) = \int_{\partial Q} d\ell = \frac{1}{2}\langle \delta_D(\alpha-\nu)\,|\eta_1| \rangle \sim N_1(\nu)$$



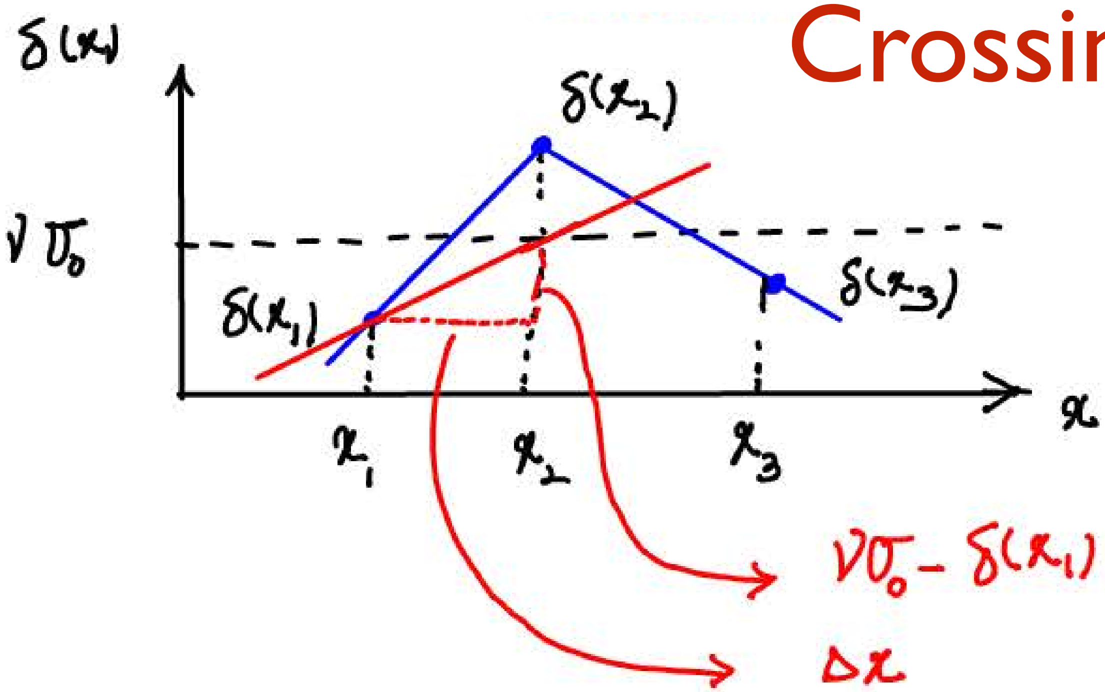
2D field

- $V_0(\nu) = \int_Q dA = \langle \Theta(\alpha-\nu) \rangle$
- $V_1(\nu) = \frac{1}{4}\int_{\partial Q} d\ell = \frac{\pi}{8}\langle \delta_D(\alpha-\nu)\,|\eta_1| \rangle \sim N_1(\nu)$

$$V_2(\nu) = \frac{1}{2\pi}\int_{\partial Q} \kappa\, d\ell = -\frac{1}{2}\langle \delta_D(\alpha-\nu)\,\delta_D(\eta_1)\,|\eta_2|\,\xi_{11} \rangle$$



Crossing from Mathematics



① $\delta(x_1) < \nu\sigma_0$

② $\dfrac{\nu\sigma_0 - \delta(x_1)}{\Delta x} < \eta_x \Rightarrow \delta(x_1) > \nu\sigma_0 - |\eta_x| \Delta x$

$$\Delta x\, N_1(\nu) = \lim_{\Delta x \to 0} \int d\eta_x \int_{\nu\sigma_0 - |\eta_x|}^{\nu\sigma_0} d\delta \; P(\eta_x, \delta)$$

1D= no. of crossing

$$= \int d\eta_x \; \Delta x\, |\eta_x|\, P(\eta_x, \delta = \nu\sigma_0) = \left\langle \delta_D(\alpha - \nu)\, |\eta_1|\, \Theta(\eta_1) \right\rangle$$

2D= mean length of iso-density contour

$$N_2(\nu) = \int d\eta_x\, d\eta_y \left(\eta_x^2 + \eta_y^2\right)^{1/2} P(\vec{\eta}, \nu\sigma_0) = \left\langle \delta_D(\alpha - \nu)\, |\eta_x^2 + \eta_y^2|^{1/2} \right\rangle$$

3D= mean surface of iso-density region

$$N_3(\nu) = \int d\eta_x\, d\eta_y\, d\eta_z\, |\vec{\eta}|\, P(\vec{\eta}, \nu\sigma_0) = \left\langle \delta_D(\alpha - \nu)\, |\eta_x^2 + \eta_y^2 + \eta_z^2|^{1/2} \right\rangle$$

Theoretical approach for Number density of Up-crossing

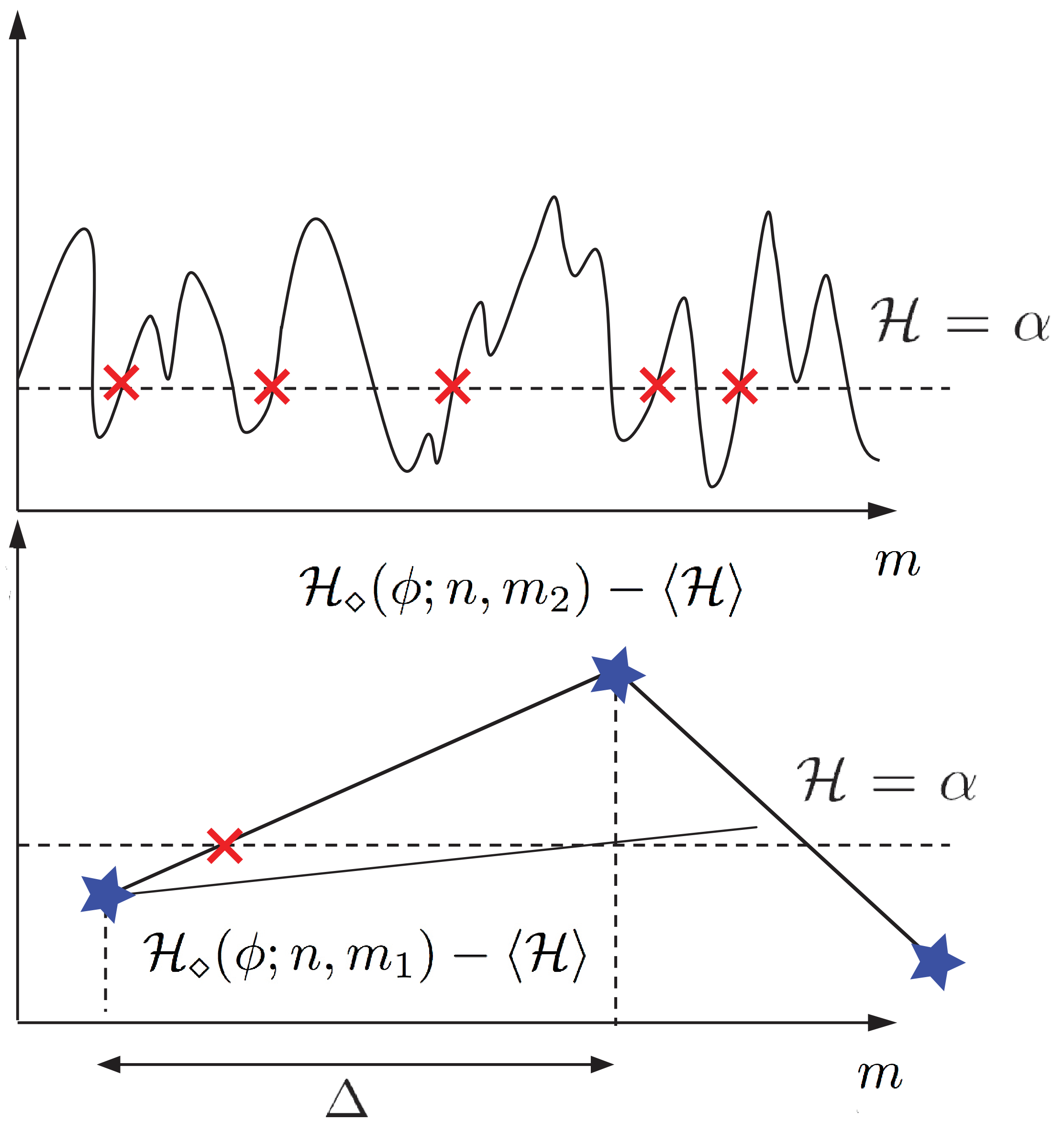
$$\langle n_p(\vec{r}) \rangle = \sum_{\vec{r}_p} \delta_D(\vec{r}-\vec{r}_p) = \int dA^\mu P(A^\mu)\, \delta_D(\vec{r}-\vec{r}_p)$$

$$A^\mu : \left(\alpha(\vec{r}), \vec{\eta}(\vec{r}), \xi_{ij} \right)$$

$$\langle n(\vec{r}) \rangle = \int dA^\mu \left[\text{Transfer function} \right] P(A^\mu)$$

Ex 2: Up-Crossing for $\alpha(\vec{r}) = \nu\sigma_0$

$$\langle n_u(\vec{r}) \rangle = \langle \delta_D(\alpha - \nu\sigma_0)\, |\vec{\eta}|\, \Theta(\vec{\eta}) \rangle$$



Perturbative parts in D-dimension for Isotropic field

$$N_1(\nu) = \langle \delta_D(\alpha-\nu)|\eta_1|\,\Theta(\eta_1)\rangle \sim [\text{Independent parameter}]^{-1}$$

$$= N_1^G(\nu) + \text{Perturbative Parts}$$

$$= \frac{1}{\pi}\frac{\sigma_1}{\sqrt{D}\,\sigma_0} e^{-\nu^2/2} + N_1^{NG}(\nu)$$

$$= N_1^G(\nu)\left[1 + A\sigma_0 + B\sigma_0^2 + \mathcal{O}(\sigma_0^3)\right]$$

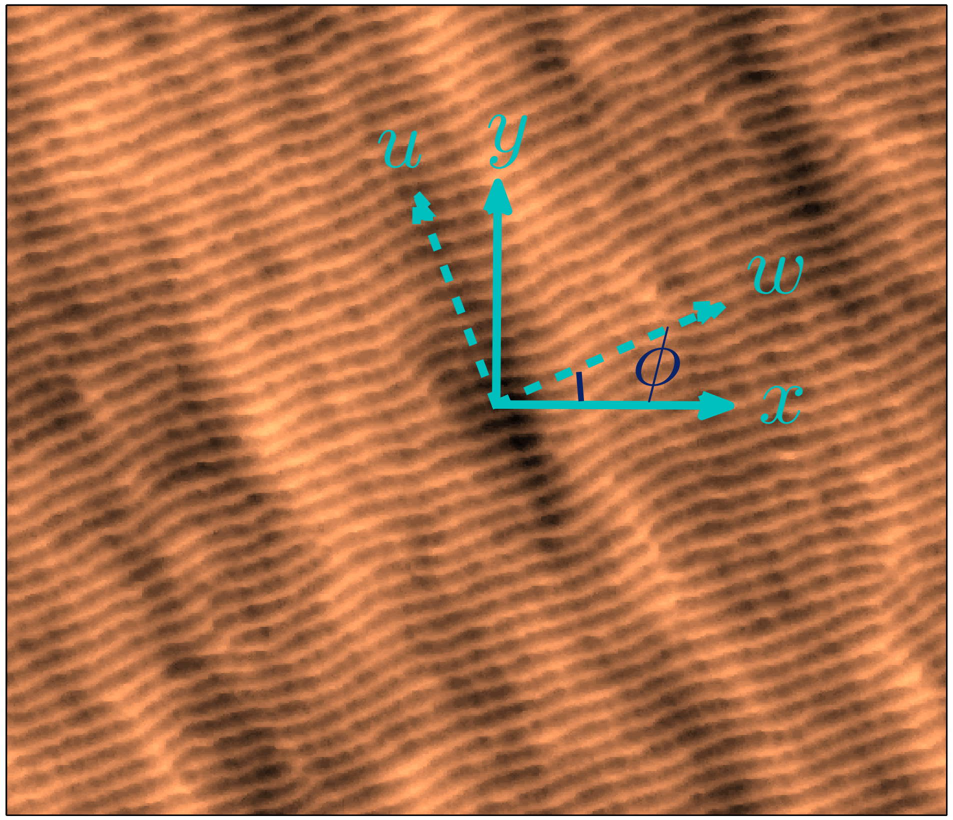
$$A = \frac{S}{6}H_3(\nu) + \frac{S^{(1)}}{3}H_1(\nu)$$

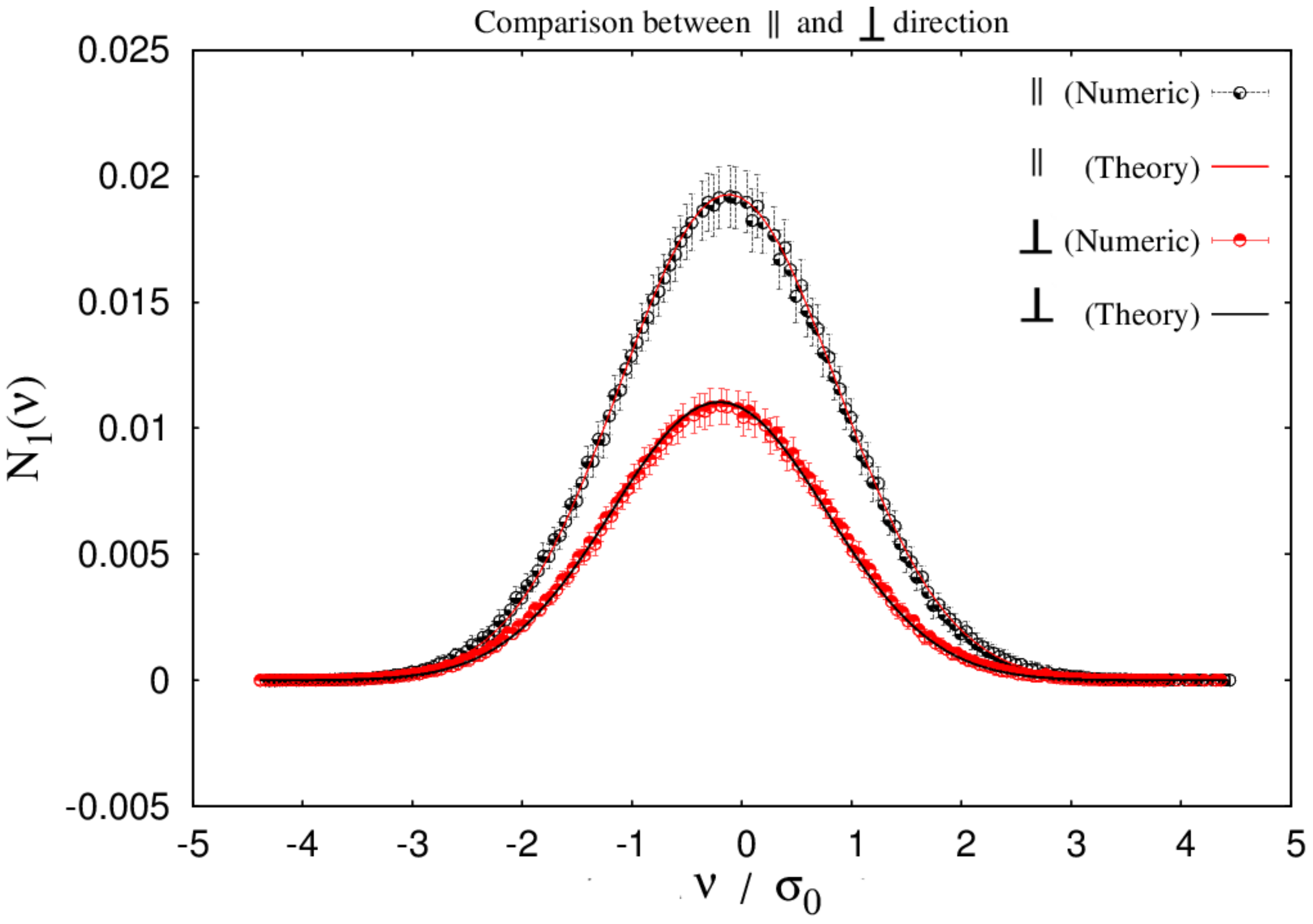
$$B = \frac{1}{24}(K - SS^{(1)})H_4(\nu) - H_2(\nu)\left(\frac{1}{12}K^{(1)} + \frac{1}{96}S^{(1)2}\right) + \frac{1}{72}S^2 H_6(\nu) + \frac{1}{8}(-K^{(3)})$$

$$S = \frac{\langle\alpha^3\rangle}{\sigma_0}, \quad S^{(1)} = -\frac{3}{4}\frac{\langle\alpha^2\nabla^2\alpha\rangle}{\sigma_0^2\sigma_1^2}, \quad K^{(3)} = \frac{\langle\nabla f^4\rangle}{2\sigma_0^2\sigma_1^4}$$

S.M.S.M et. al., In progress

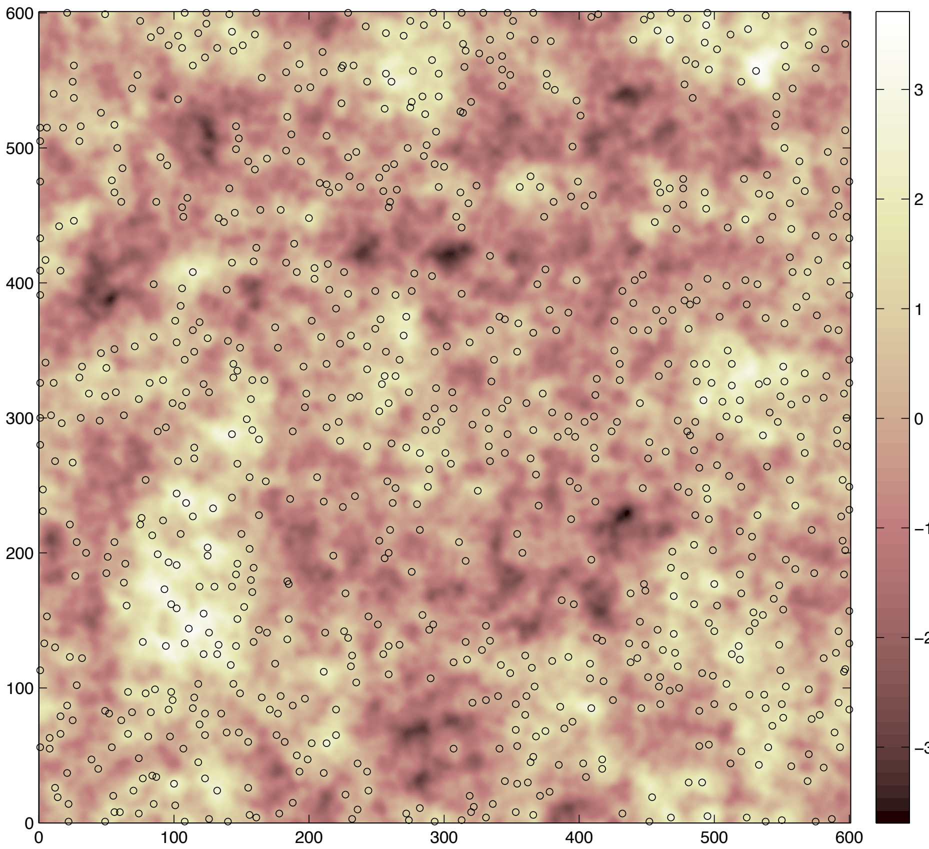
Application for an anisotropic 2+1 D



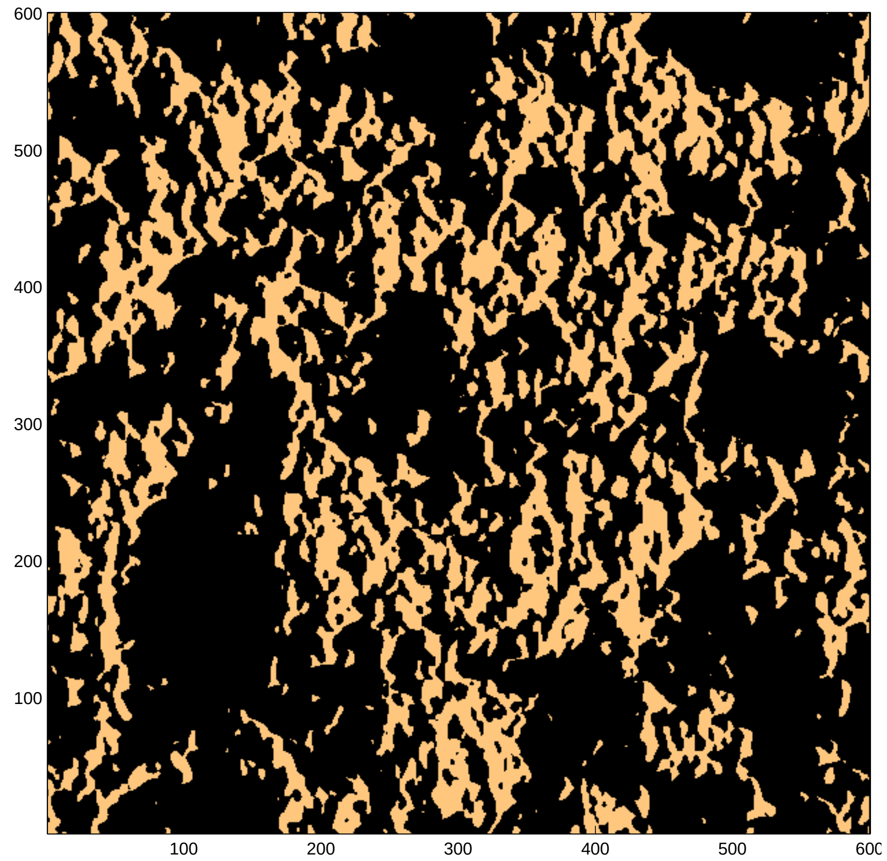


M. Ghasemi Nezhadhaghighi, S. M. S. M. T. Yasseri, S. M. Vaez Allaei,, Journal of Applied Physics, 2015

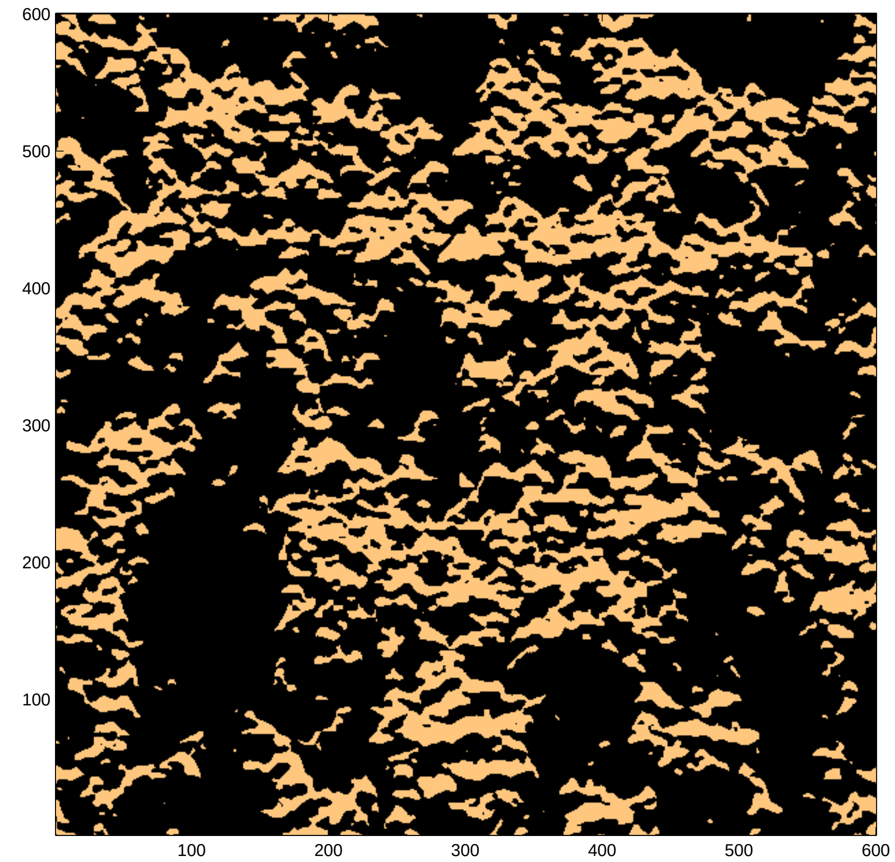
Clustering of Up-Crossing



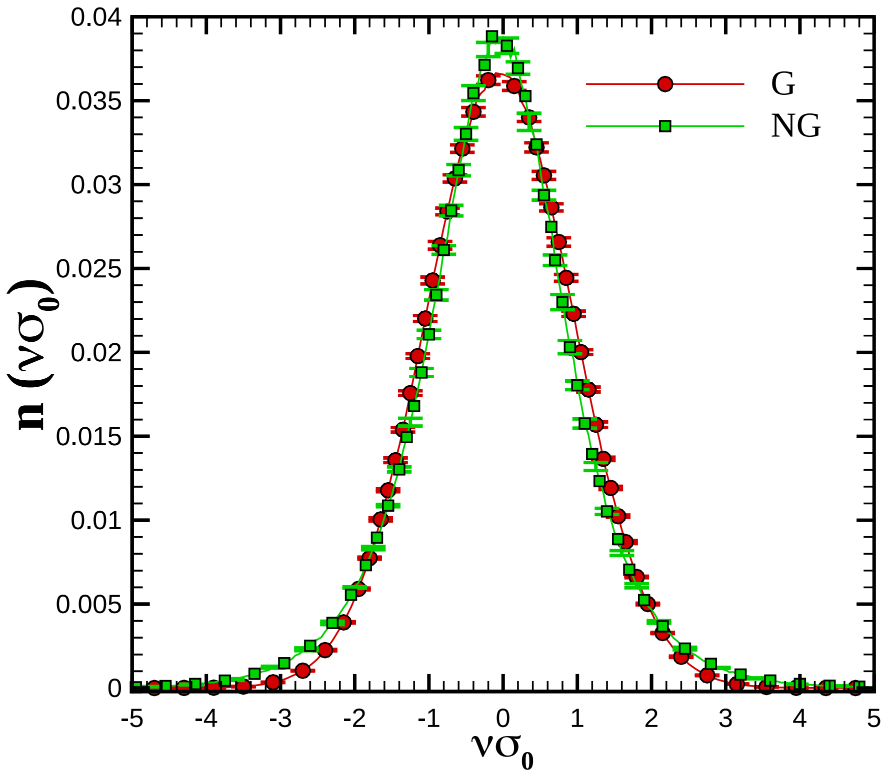
Gaussian field with peaks

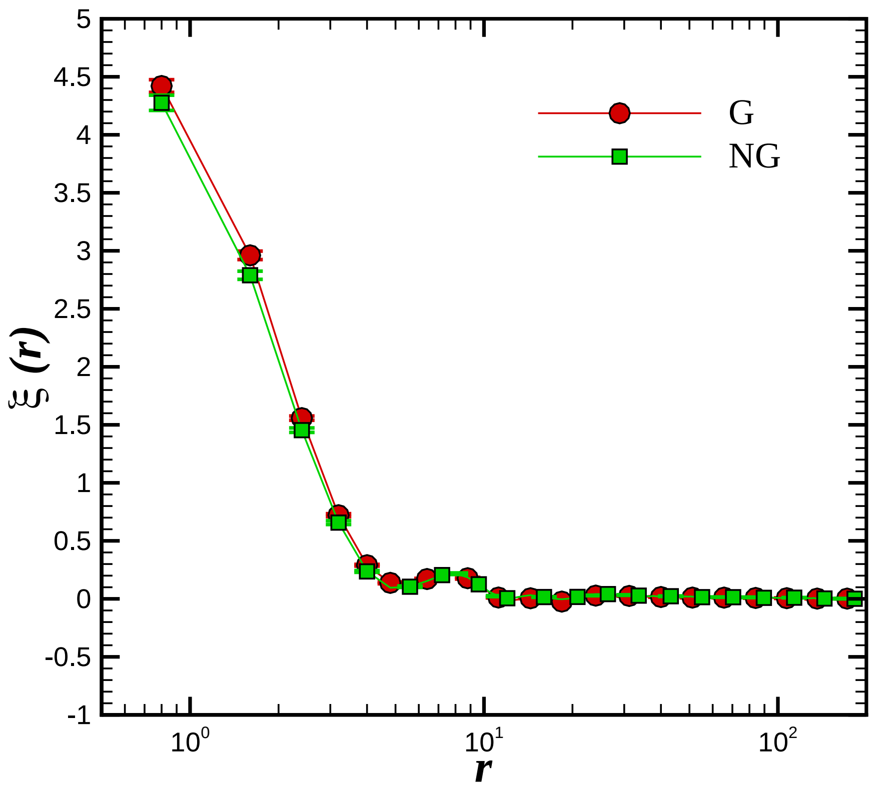


crossing in y above a threshold



crossing in x above a threshold

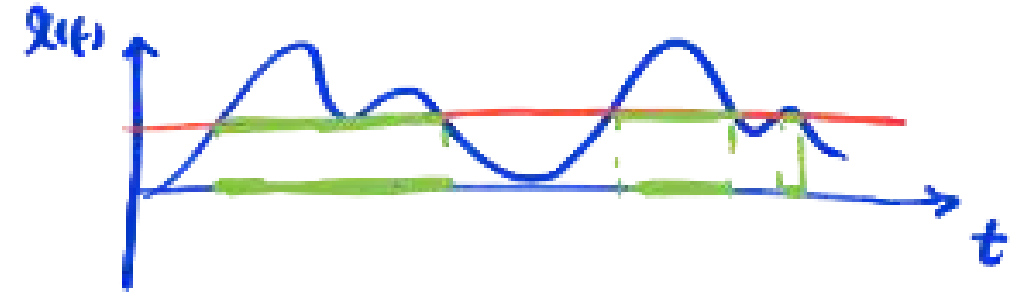




Import Data $\{x_i\}$. $i = 1, \dots N$

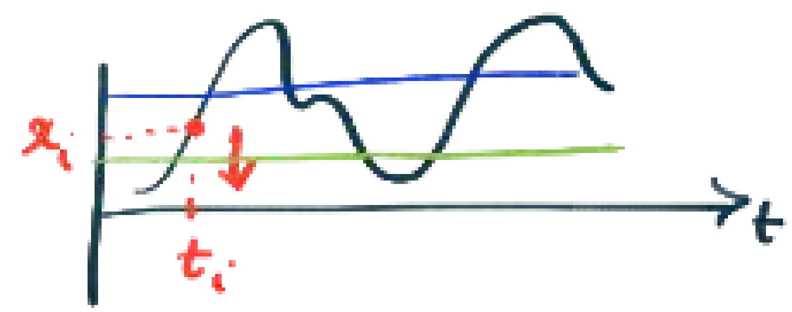
$$\Delta x = \frac{Max(x) - Min(x)}{M}$$

Minkowski 0



loop on Data $i = 1, N$

$$K = \frac{x(i)}{\Delta x}$$



loop $l = K_{min}, K, \Delta K$

$N0(l) = N0(l) + 1$

End loop

تمام ترازهای پایین‌تر را تحت
تأثیر قرار می‌دهد.

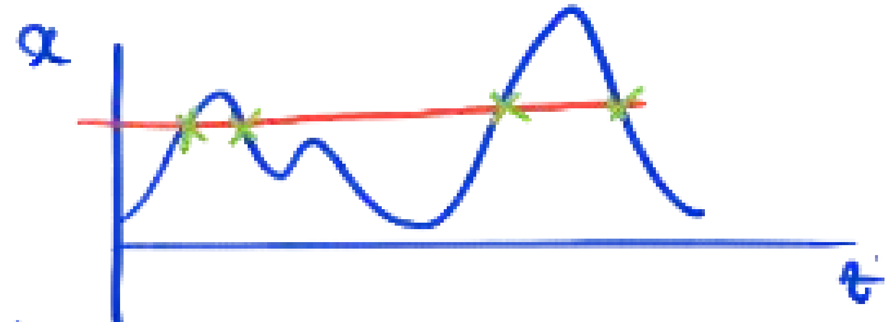
در تراز سبز سهم دارد ولی در تراز آبی [illegible]

End loop

$$N0 = \frac{N0}{N}$$

بهنجار می‌شود

Minkowski 1



loop on Data $i = 1, N-1$

$K_1 = \frac{x(i)}{\Delta x}$

$K_2 = \frac{x(i+1)}{\Delta x}$

If $x(i+1) \geq x(i)$

loop $l = K_1, K_2, \Delta K$

$N1(l) = N1(l) + 1$

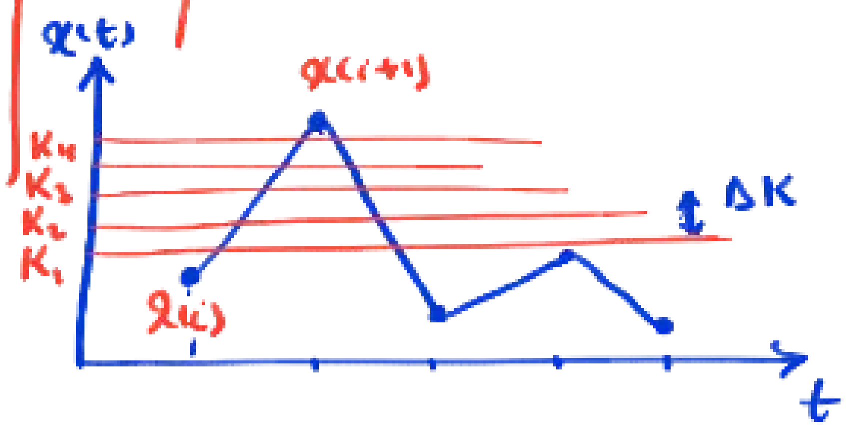
End loop

End if

End loop

فقط برخوردها با شیب مثبت را شمرده‌ایم

ترازهای مختلف



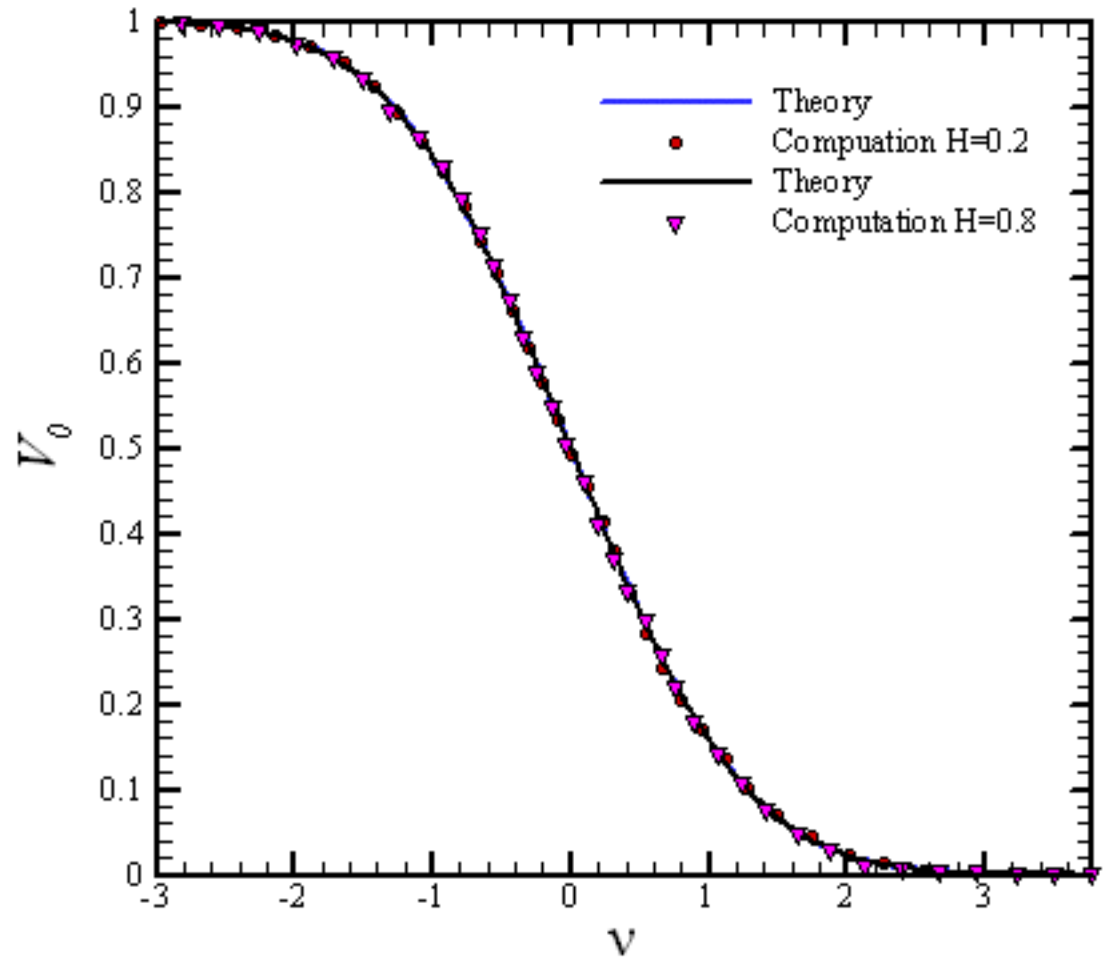
$N1 = N1/N$

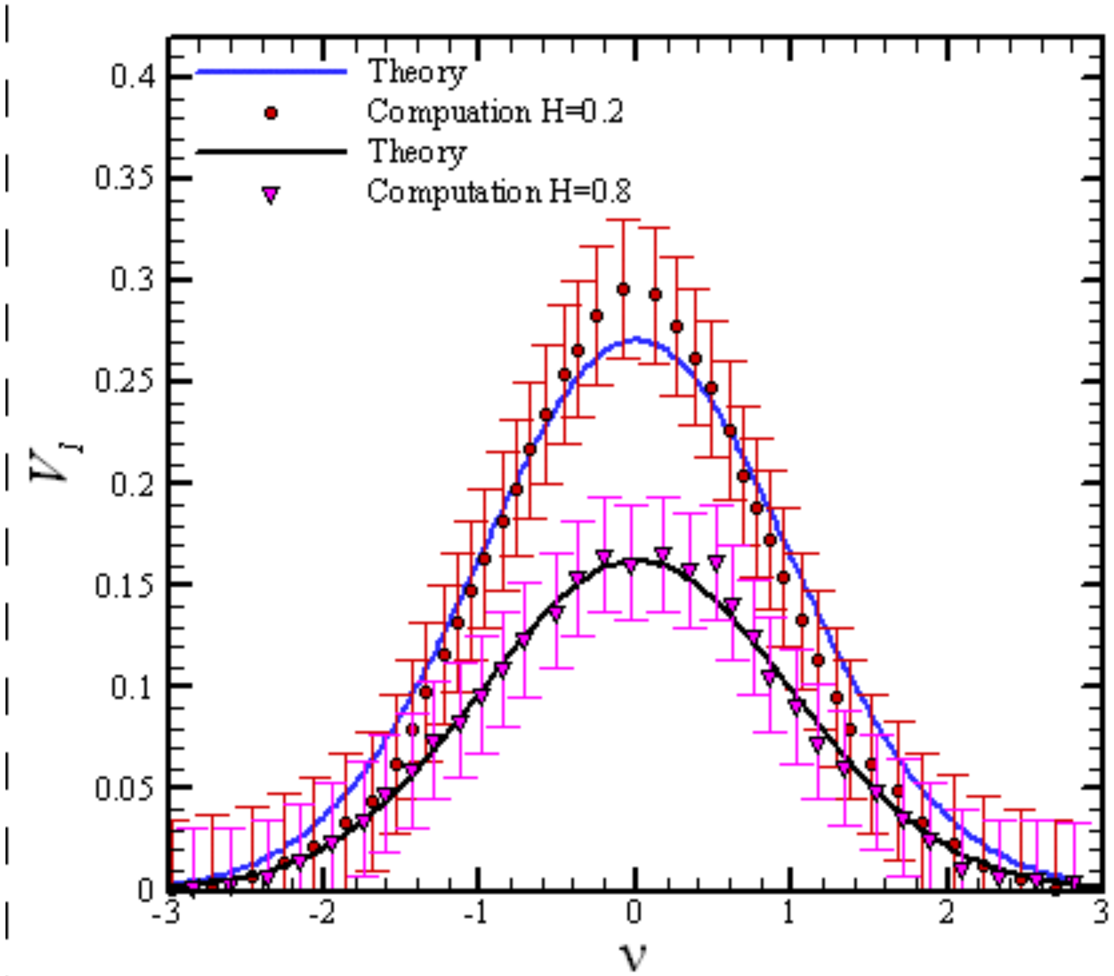
$N_1 = 2N_1$ ← only for Stationary and Regular Data

End Program

تمرین

نتایج برای داده های مشخص شده





Some advantages of Crossing statistics

A) Simple for implementation.
B) Directional nature
C) Determining the kinds of anisotropies
D) More sensitive to find exotic feature

از توجه شما سپاسگزارم