Coherent states, Support Vector Machines and function estimation

Michaël Fanuel

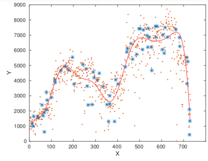
November 18, 2016

KULeuven, Belgium
ESAT-STADIUS, Group of J.A.K. Suykens

Outline

Motivations

Motivations: Finite dimensional LS-SVM regression



Given a sample $\{(x_i, y_i)\}_{i=1}^N$ where $(x_i, y_i) \in \mathbb{R}^d \times \mathbb{R}$
Minimization of the errors with a regularization term

$$\min_{\substack{w \in \mathbb{R}^h \\ e_i, b \in \mathbb{R}}} \frac{1}{2}\|w\|_2^2 + \frac{\gamma}{2}\sum_{i=1}^N e_i^2 \quad \text{s.t.} \quad y_i = w^T\varphi(x_i) + b + e_i,$$

with $i = 1, \ldots, N$, $\gamma > 0$ and where $\varphi : \mathbb{R}^d \to \mathbb{R}^h$. The solution is obtained thanks to the symmetric positive definite kernel

$$K(x, x') = \varphi^T(x)\varphi(x') \qquad \text{for } x, x' \in \mathbb{R}^d$$

The Lagrange multipliers are obtained by solving a linear system

$$y_i \quad = \quad \sum_{j=1}^N \alpha_j K(x_j, x_i) + b + e_i,$$

for the Lagrange multipliers α's. The prediction is done by the "model":
$y(x) = \sum_{j=1}^N \alpha_j K(x_j, x) + b$.

Extension to infinite dimensional Hilbert space

Usually, in machine learning, use of the popular gaussian $K_\sigma(x, x') = \exp(-\frac{1}{2\sigma^2}\|x - x'\|_2^2)$.

- Infinite dimensional feature space?
  Non rigorous view:

$$x \quad \mapsto \quad \varphi(x) = \begin{pmatrix} \varphi_1(x) \\ \varphi_2(x) \\ \vdots \end{pmatrix},$$

- Generalization?

$$\begin{array}{rcl} \mathbb{R}^d & \to & \mathcal{H} \\ x & \mapsto & |\eta_x\rangle, \end{array}$$

$\mathcal{H}$ is an infinite dimensional Hilbert (separable), with inner product $\langle\cdot|\cdot\rangle_{\mathcal{H}}$.

Light introduction: Let us assume

- RKHS $\{\mathcal{H}_K, \langle\cdot|\cdot\rangle_K\}$ orthonormal basis $\{\psi_n\}_{n=0}^{\infty}$ with $\sum_{n=0}^{+\infty} |\psi_n(x)|^2 < \infty$ for all $x \in X$.

[1] A. Horzela and F. H. Szafraniec J. Phys. A45 (24),244018,2012

Coherent states from a Reproducing Kernel Hilbert Space

Light introduction: Let us assume

- RKHS $\{\mathcal{H}_K, \langle\cdot|\cdot\rangle_K\}$ orthonormal basis $\{\psi_n\}_{n=0}^{\infty}$ with $\sum_{n=0}^{+\infty} |\psi_n(x)|^2 < \infty$ for all $x \in X$.
- Hilbert space $\{\mathcal{H}, \langle\cdot|\cdot\rangle_{\mathcal{H}}\}$ orthonormal basis $\{\phi_n\}_{n=0}^{\infty}$.

[1] A. Horzela and F. H. Szafraniec J. Phys. A45 (24),244018,2012

Coherent states from a Reproducing Kernel Hilbert Space

Light introduction: Let us assume

- RKHS $\{\mathcal{H}_K, \langle\cdot|\cdot\rangle_K\}$ orthonormal basis $\{\psi_n\}_{n=0}^{\infty}$ with $\sum_{n=0}^{+\infty} |\psi_n(x)|^2 < \infty$ for all $x \in X$.
- Hilbert space $\{\mathcal{H}, \langle\cdot|\cdot\rangle_{\mathcal{H}}\}$ orthonormal basis $\{\phi_n\}_{n=0}^{\infty}$.

A "measure-free" coherent state (non-normalized)[1]

$$|\eta_x\rangle = \sum_{n=0}^{\infty} \bar{\psi_n}(x)|\phi_n\rangle \in \mathcal{H}.$$

Reproducing kernel of $\mathcal{H}_K$ is $K(x,x') = \sum_{n=0}^{\infty} \psi_n(x)\bar{\psi_n}(x') = \langle\eta_x|\eta_{x'}\rangle_{\mathcal{H}}$.

[1] A. Horzela and F. H. Szafraniec J. Phys. A45 (24),244018,2012

Coherent states and SVM

Primal problem

$$\min_{\substack{|w\rangle\in\mathcal{H}\\ e_i,b\in\mathbb{R}}} \frac{1}{2}\langle w|w\rangle_{\mathcal{H}} + \frac{\gamma}{2}\sum_{i=1}^{N} e_i^2 \qquad \text{subject to} \quad y_i = \langle w|\eta_{x_i}\rangle_{\mathcal{H}} + b + e_i,$$

for $i = 1, \dots, N$. The solution is obtained by solving a linear system

$$y_i \quad = \quad \sum_{j=1}^{N} \alpha_j \langle \eta_{x_j}|\eta_{x_i}\rangle_{\mathcal{H}} + b + e_i,$$

Kernel function $K(x_j, x_i) = \langle \eta_{x_j}|\eta_{x_i}\rangle_{\mathcal{H}}$,

Model used for prediction:

$$y(x) = \sum_{j=1}^{N} \alpha_j K(x_j, x) + b.$$

Design of a kernel

In machine learning, the self-tuned kernel[2] is used

$$K^s(x,x') = \exp\Big(-\frac{\|x-x'\|_2^2}{\sigma_x\sigma_{x'}}\Big).$$

One alternative possibility is to associate the pair $(x,\sigma_x) \mapsto \psi_{(x,\sigma_x)}(\cdot) = \psi(\frac{\cdot - x}{\sigma_x})$ so that the kernel

$$k(x,x') = \langle \psi_{(x,\sigma_x)} | \psi_{(x',\sigma_{x'})} \rangle$$

can be used, if there is an empirical way to estimate σ_x (related statistical theory).

[2] T. Berry, J. Harlim, Variable bandwidth diffusion kernels, Applied Comput. Harmon. Anal., 40(1),2016 (in the context of diffusion maps)

Wavelets on graphs

Broader context

- Graph-based data analysis methods
  Diffusion maps[3] Diffusion wavelets[4]
- Signal processing on graphs
  Spectral Graph Wavelet[5]
  Signal processing on graphs[6]
  Compressive spectral clustering[7]
  Decomposition of pictures in patches...
- Common basic ingredient: the combinatorial Laplacian on graphs.

---

[3] R.R. Coifman, S. Lafon, Applied Comput. Harmon. Anal., 21(1), 5-30, 2006
[4] R.R. Coifman, M. Maggioni, Applied Comput. Harmon. Anal., 21(1), 53-94, 2006
[5] J.-P. Antoine *et al.* Applied Comput. Harmon. Anal., 28(2):189 - 202, 2010
D. K. Hammond *et al.* Applied Comput. Harmon. Anal., 30(2):129 - 150, 2011
[6] D. I. Shuman *et al.* IEEE Signal Processing Magazine,30(3), 83-98, 2013
[7] N. Tremblay, G. Puy, R. Gribonval, P. Vandergheynst, In ICML, June 2016

Discrete differential operators

Definition of the wavelets in Fourier space. Connected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with a finite set of nodes $\mathcal{V}$ and $\mathcal{E}$ edges with positive weights $w_{ij} = w_{ji} \geq 0$. Given a function on the nodes $p(i)$, its gradient on the edge $[i, j]$ reads

$$\mathrm{d}p(i, j) = p(j) - p(i).$$

A (discrete) vector field is given by a skew symmetric matrix element $a_{ij} = -a_{ji}$ mapped to each pair of nodes i and j connected by an edge. The divergence of the vector field is simply

$$\text{div } a(i) = \sum_{j \in \mathcal{V}} w_{ij} a_{ij},$$

which corresponds to the adjoint of the discrete gradient: $-\mathrm{d}^* a$, with respect to the inner products:

$$\langle p, p' \rangle = \sum_{i \in \mathcal{V}} p(i) p'(i), \quad \text{and} \quad \langle a, a' \rangle_{\mathcal{E}} = \frac{1}{2} \sum_{[i,j] \in \mathcal{E}} w_{ij} a(i, j) a'(i, j).$$

Discrete Laplacian

Composing the divergence of the gradient, we obtain the combinatorial Laplacian

$$-(L_0 p)(i) = \sum_{j \in \mathcal{V}} w_{ij} \Big(p(j) - p(i) \Big),$$

defined as a *positive semi-definite* operator (convention). Its eigenvalues are $0 = \lambda_0 < \lambda_1 \leq \cdots \leq \lambda_{N-1}$, and the eigenvectors satisfy

$$L_0 |u_\ell\rangle = \lambda_\ell |u_\ell\rangle.$$

Notice that if $|\delta_i\rangle$ is a discrete Dirac, then $p(i) = \langle \delta_i | p \rangle$. The u_ℓ's are the discrete Fourier modes of the graph. Notice that $u_0 = \mathrm{cst}$. In particular we can have a diffusion equation

$$\frac{\mathrm{d}p_t(i)}{\mathrm{d}t} = -\Big(L_0 p_t \Big)(i),$$

solved by using the propagator $T^t = \exp(-tL_0)$ (or heat kernel).

Broader context

Deformed combinatorial Laplacians

- Connection Laplacian [8](Synchronization of rotations)
- Magnetic Laplacian [9]
- Signed Laplacian [10](Clustering signed graphs)
- Dilation Laplacian[11](Ranking objects from pairwise comparisons)

---

[8]A. Singer and H-T. Wu, Vector Diffusion Maps and the Connection Laplacian, Comm. Pure Appl. Math., 65: 1067-1144, 2012

[9]M. A. Shubin. Discrete magnetic Laplacian, Comm. Math. Phys.,164, 259-275, 1994

[10]J. Kunegis *et al.* Spectral Analysis of Signed Graphs for Clustering, Prediction and Visualization, chapter 48, 559-570, 2010

[11]M. Fanuel and J.A.K. Suykens, Deformed Laplacians and spectral ranking in directed networks, arxiv:1511.00492

Towards graph wavelets

- In general, we have no translation group, no rotation group. We use only the existence of a “Fourier Space”.
- Two main constructions exist: the diffusion wavelets (orthogonal) and the spectral graph wavelets.
- In the diffusion wavelets case, we assume a semi-group $\{T^t\}_{t\geq 0}$ associated to a diffusion operator (for instance, $T = \exp(-\epsilon\Delta)$).
- For the spectral graph wavelets case, the construction is based on an analogy with 1-D wavelet in Fourier space

$$(T^s f)(x) = \int_{-\infty}^{+\infty} \mathrm{d}y \frac{1}{s}\bar{\psi}(\frac{y-x}{s}) f(y) = \int_{-\infty}^{+\infty} \frac{\mathrm{d}k}{2\pi} e^{\mathrm{i}kx} \bar{\hat{\psi}}(sk)\hat{f}(k).$$

Notice that $e^{\mathrm{i}kx}$ is an eigenfunction of the Laplacian $\mathrm{d}^2/\mathrm{d}x^2$.

Spectral Graph Wavelets

Defined in Fourier Space[12]. Given $g : \mathbb{R}^+ \to \mathbb{R}$ satisfying $g(0) = 0$ and the admissibility condition

$$\int_0^{+\infty} \frac{\mathrm{d}t}{t} g^2(t) = c < \infty,$$

a spectral graph wavelet at i and of scale parameter $t > 0$ is

$$|\psi_{i,t}\rangle = g(-tL_0)|\delta_i\rangle = \sum_{\ell=1}^{N-1} g(-t\lambda_\ell) u_\ell(i) |u_\ell\rangle.$$

Resolution of the identity

$$\frac{1}{c} \sum_{i \in \mathcal{V}} \int_0^{+\infty} \frac{\mathrm{d}t}{t} |\psi_{i,t}\rangle\langle\psi_{i,t}| = \mathbb{I} - |u_0\rangle\langle u_0|,$$

We observe that they are symmetric waveforms.

[12] J.-P. Antoine *et al.* Applied Comput. Harmon. Anal., 28(2):189 - 202, 2010
D. K. Hammond *et al.* Applied Comput. Harmon. Anal., 30(2):129 - 150, 2011

Diffusion with drift

- Starting point: Fokker-Planck equation

$$\frac{\mathrm{d}p_t}{\mathrm{d}t}(x) = -\big(\nabla \cdot \vec{J}_t\big)(x), \text{ with } \vec{J}_t(x) = -\vec{\nabla} p_t(x) + \vec{\mu}(x) p_t(x).$$

- Discrete analogue? Fix a discrete vector field $a_{ij} = -a_{ji}$ and define the rate $r^{(\beta)}_{j \to i} = \exp(\beta a_{ji}/2)$.

$$\frac{\mathrm{d}p_t(i)}{\mathrm{d}t} = -\big(\mathrm{div} J_t\big)(i), \text{ with } J_t(i,j) = -\Big(r_{j\to i} p_t(j) - r_{i \to j} p_t(i)\Big).$$

We have

$$J_t(i,j) = -\mathrm{d}p_t(i,j) + \beta a_{ij} \frac{1}{2}\Big(p_t(i) + p_t(j)\Big) + \mathcal{O}(\beta^2).$$

This provides a deformed "Laplacian" L_β given by the rhs of

$$\frac{\mathrm{d}}{\mathrm{d}t} p_t(i) = \sum_{j \in \mathcal{V}} w_{ij} \Big(p_t(j) e^{\beta a_{ji}/2} - p_t(i) e^{\beta a_{ij}/2}\Big) \triangleq -(L_\beta p_t)(i).$$

Deformed gradient

The current is given by

$$J_t(i,j) = -\Big(\exp(\beta a_{ji}/2)p_t(j) - \exp(\beta a_{ij}/2)p_t(i)\Big) = -(\mathrm{d}_a p_t)(i,j)$$

"Covariance property"

$$\Big(\mathrm{d}_a e^{\alpha} p\Big)(i,j) = e^{\frac{\alpha(i)+\alpha(j)}{2}}\Big(\mathrm{d}_{a-\mathrm{d}\alpha} p\Big)(i,j)$$

which is a discrete analogue of

$$-\vec{\nabla}(e^{\alpha}p)(x) + \vec{\mu}(x)(e^{\alpha}p)(x) = e^{\alpha(x)}\Big(-\vec{\nabla}p(x) + \big(\vec{\mu}(x) - \vec{\nabla}\alpha(x)\big)p(x)\Big).$$

See also[13].

[13] R. Kenyon, Ann. Probab., 39, 5 (2011)

Diffusion with drift

The properties of the deformed Laplacian

$$L_\beta p(i) = \sum_{j \in \mathcal{V}} w_{ij} \left(p(i) e^{\beta a_{ij}/2} - p(j) e^{\beta a_{ji}/2} \right)$$

are:

- It is non-symmetric, but the real part of its eigenvalues is positive.
- There is a stationary distribution m_β such that $L_\beta m_\beta = 0$.
- By construction, $\sum_{i \in \mathcal{V}} L_\beta p(i) = 0$.

Diffusion with drift

Connection with a random walk:

$$L_\beta p(i) = \sum_j p(j) r_j \Big(\delta_{i,j} - P_\beta(j \to i)\Big),$$

where $r_j = \sum_{k\in\mathcal{V}} w_{jk} r^{(\beta)}_{j\to k}$ and the transition matrix

$$P_\beta(j \to i) = \frac{w_{ji} r^{(\beta)}_{j\to i}}{\sum_{k\in\mathcal{V}} w_{jk} r^{(\beta)}_{j\to k}} \text{ with } r^{(\beta)}_{j\to i} = e^{\beta a_{ji}/2}.$$

There exists a stationary distribution $\sum_j \pi(j) P_\beta(j \to i) = \pi(i)$. Then, we have

$$L_\beta m_\beta = 0$$

with

$$m_\beta(j) = \frac{\pi(j)/r_j}{\sum_{k\in\mathcal{V}} \pi(k)/r_k} > 0.$$

The eigenvalues of L_β depend on a_{ij}. More precisely, we have a Hodge decomposition

$$a = \mathrm{d}h + a_M.$$

We ask a detailed balance condition (Kolmogorov criterion), then $a_{ij} = -\mathrm{d}U(i,j)$.

Generator of the diffusion with drift

If $a_{ij} = -\mathrm{d}U(i,j)$, then the stationary distribution is the Boltzmann-Gibbs measure

$$\mu_\beta(i) = \frac{1}{Z_\beta} e^{-\beta U(i)}, \text{ with } Z_\beta = \sum_{i \in \mathcal{V}} e^{-\beta U(i)},$$

and $L_\beta^\dagger$ is quasi-Hermitian[14] (pseudo-Hermitian[15]), that is,

$$\hat{\mu}_\beta L_\beta^\dagger = L_\beta \hat{\mu}_\beta .$$

with the Hermitian strictly positive operator $\hat{\mu}_\beta$ with matrix elements $\mu_\beta(i,j) = \mu_\beta(i)\delta_{i,j}$.

[14] J.-P. Antoine and C. Trapani, J. Math. Phys. 55, 013503 (2014)
[15] A. Mostafazadeh, J. Math. Phys. 43, 205-214 (2002)

Generator of the diffusion with drift

The generator L_β has non-negative eigenvalues $0 = \lambda_0 < \lambda_1 \leq \cdots \leq \lambda_{N-1}$ associated to real right-eigenvectors v_ℓ satisfying the orthonormality condition

$$\langle v_\ell | w_{\ell'} \rangle = \delta_{\ell,\ell'}$$

and we have $v_0 = \mu_\beta$. Spectral representation

$$L_\beta = \sum_{\ell=0}^{N-1} \lambda_\ell |v_\ell\rangle\langle w_\ell|$$

with bi-orthonormal basis of eigenvectors. More interestingly, we have the resolution of the identity

$$\sum_{\ell=0}^{N-1} |v_\ell\rangle\langle w_\ell| = \mathbb{I}.$$

Graph wavelets drift

Let us choose two functions[16] $g_1 : \mathbb{R}^+ \to \mathbb{R}$ and $g_2 : \mathbb{R}^+ \to \mathbb{R}$ such that $g_1(0)g_2(0) = 0$ and satisfying

$$\int_0^{+\infty} \frac{\mathrm{d}t}{t} g_1(t) g_2(t) = c_{12} < \infty.$$

Then, two sets of spectral graph wavelets centered at i and of scale parameter $t > 0$ are defined:

$$|\Psi_{i,t}\rangle = g_2(-tL_\beta)|\delta_i\rangle = \sum_{\ell=1}^{N-1} g_2(-t\lambda_\ell) w_\ell(i) |v_\ell\rangle.$$

and

$$|\Phi_{i,t}\rangle = g_1(-tL_\beta^\dagger)|\delta_i\rangle = \sum_{\ell=1}^{N-1} g_1(-t\lambda_\ell) v_\ell(i) |w_\ell\rangle.$$

[16] I. Daubechies. Ten Lectures on Wavelets. Society for Industrial and Applied Mathematics, 1992

Graph wavelets drift

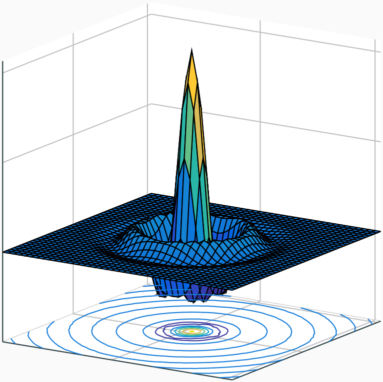
The resolution of the identity is given by

$$\frac{1}{c_{12}} \sum_{i \in \mathcal{V}} \int_0^{+\infty} \frac{\mathrm{d}t}{t} |\Psi_{i,t}\rangle\langle\Phi_{i,t}| = \mathbb{I} - |v_0\rangle\langle w_0|,$$

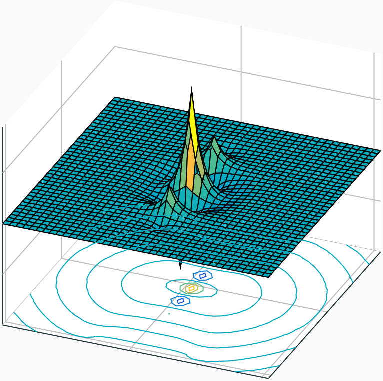
in analogy with the bi-coherent states. [17] [18](2 resolutions formulas) In the sequel, we choose $g(u) = u^2 \exp(-u^2)$.

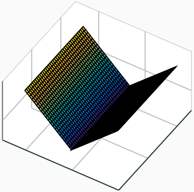
[17]J. Govaerts *et al.* J. Phys. A, 42(44):445304, 2009
[18]F. Bagarello, Geometric Methods in Physics, Trends in Mathematics pp 15-23, 2016

Illustration



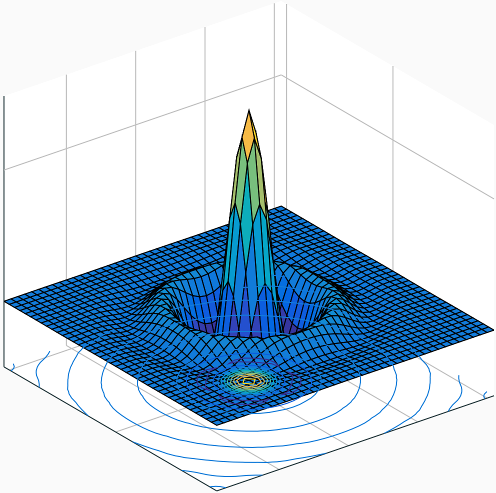
Graph wavelet with drift



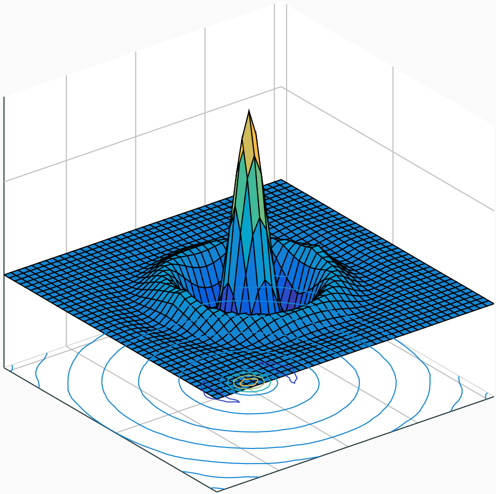


The drift is caused by a linear potential of the graph (regular grid).

Deformation of a spectral graph wavelet

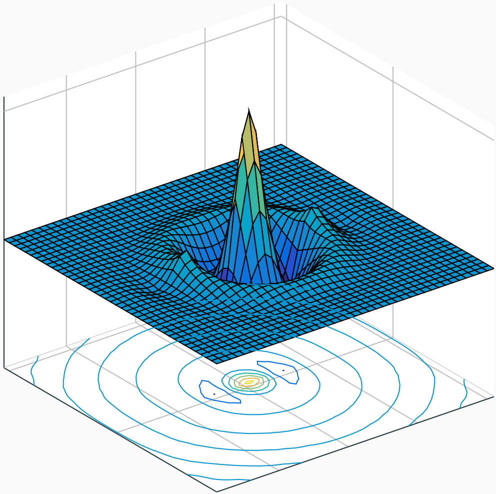


No deformation.



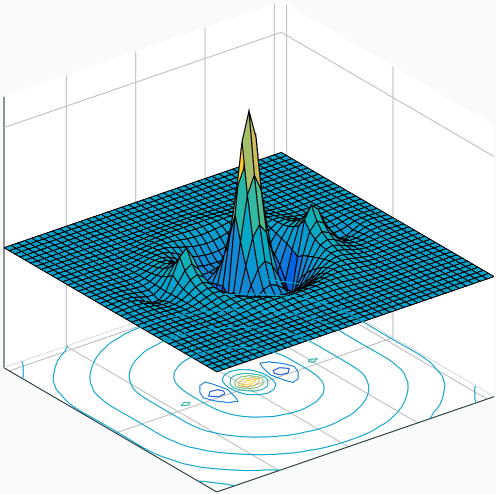
Increasing the β parameter.

Deformation of a spectral graph wavelet



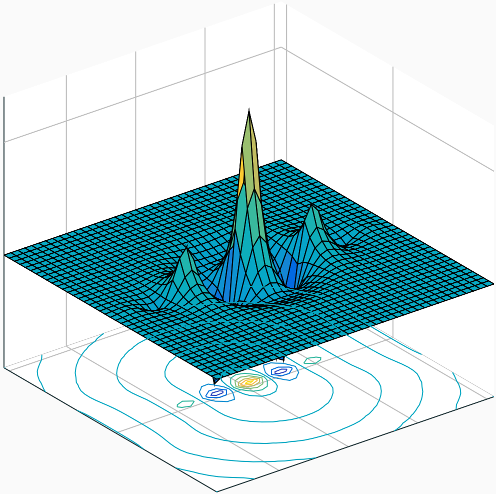
Increasing the β parameter.

Deformation of a spectral graph wavelet



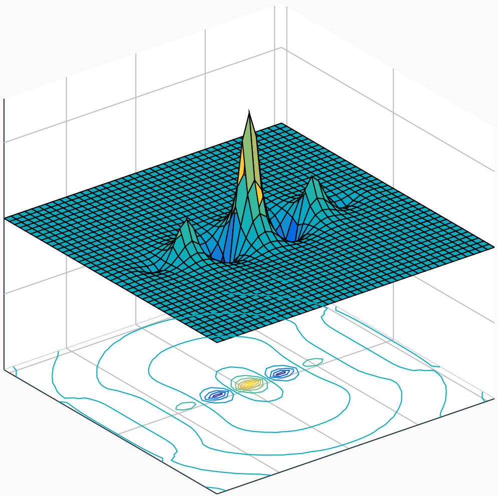
Increasing the β parameter.

Deformation of a spectral graph wavelet



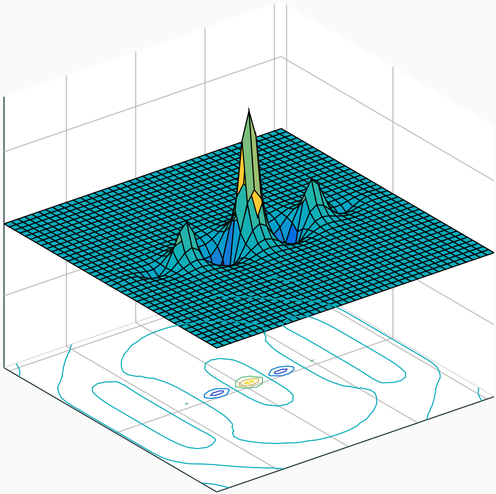
Increasing the β parameter.

Deformation of a spectral graph wavelet



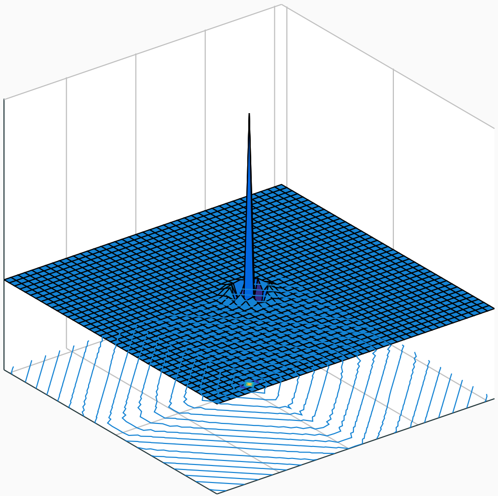
Increasing the β parameter.

Deformation of a spectral graph wavelet



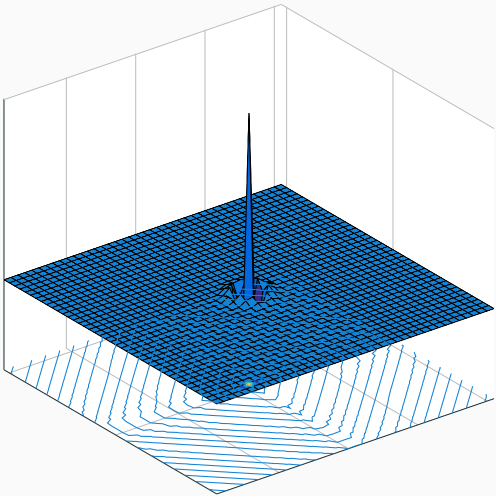
Increasing the β parameter.

Graph wavelet at various scales



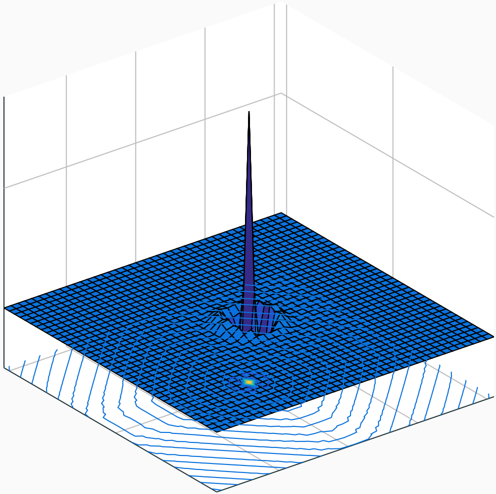
Increasing the scale parameter t.

Graph wavelet at various scales



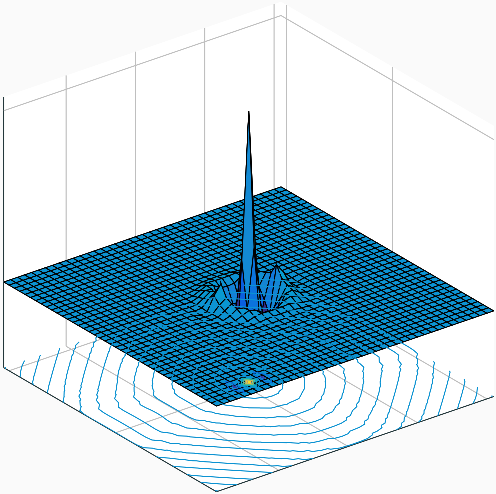
Increasing the scale parameter t.

Graph wavelet at various scales



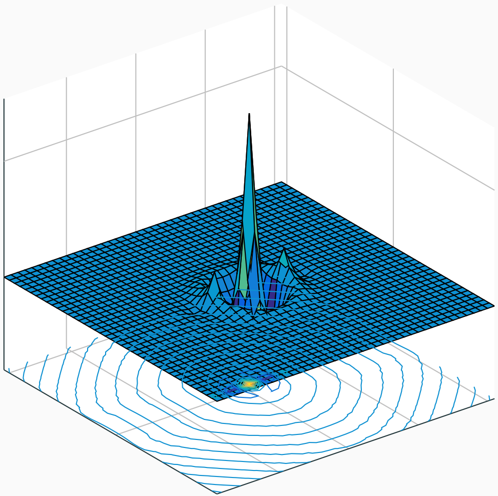
Increasing the scale parameter t.

Graph wavelet at various scales



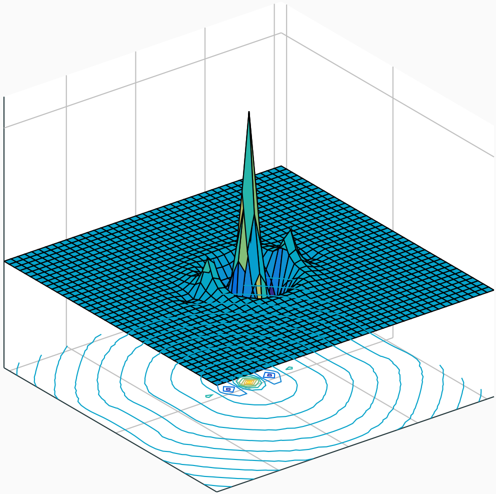
Increasing the scale parameter t.

Graph wavelet at various scales



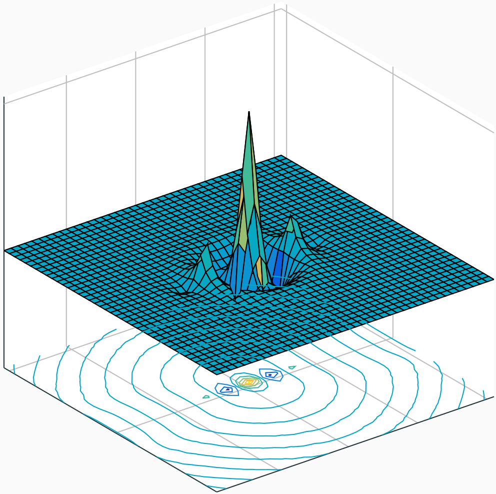
Increasing the scale parameter t.

Graph wavelet at various scales



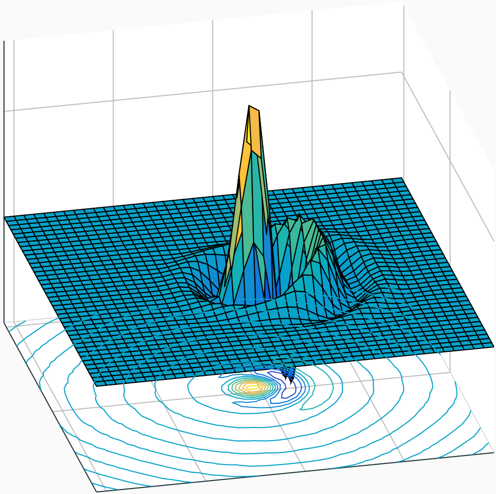
Increasing the scale parameter t.

Graph wavelet at various scales



Increasing the scale parameter t.

Drifting wavelet



Conclusions

Conclusions

- Study of directed diffusion maps.
- Design of novel graph kernels.
- Interest for signal processing?