Topic 4 Graph Denoising

Set up : $G = (V, E)$

W_{ij} $n\times n$ positive symmetric (eg. $W_{ij} = e^{-\frac{\|x_i - x_j\|^2}{\varepsilon}}$

$P = D^{-1}W$

$P = \Psi \Lambda \Phi^T$ $\quad \Psi^T D \Psi = I$ $\quad \Psi = [\psi_1, \cdots, \psi_n]$

$\Phi = D\Psi$

$f \in \mathbb{R}^n$ $\quad f: V \to \mathbb{R}$

$i \to f_i$

assumption : f is "smooth" with graph.

- Generalized Fourier coefficient $f = \sum_{k=1}^{n} c_k \psi_k$

c_k : generalized Fourier coef.

ψ_k : generalized Fourier basis

Rmk. ψ_k are the eigenfunctions of graph Laplacian.

(e^{ikx}) $\sin kx$ and $\cos kx$ are eigenfunctions of the Laplacian operator Δ.

Hence ψ_k's are called Fourier basis.

Prop. If $f(x) \in C^1[0, 2\pi]$ then $\int_0^{2\pi} |f'(x)|^2 dx \le C$.

then $\sum_{k=-\infty}^{\infty} \int_0^{2\pi} |ikc_k \cdot e^{ikx}|^2 dx$

$\le \sum_{k=-\infty}^{\infty} 2\pi |c_k|^2 \cdot k^2$ $\Rightarrow$ meaning $|c_k|$ decays faster than $\frac{1}{k}$.

Rmk. The "smoothness" (i.e. f is n-th differentiable, how large the derivatives) of f implies the decaying property of $\{c_k\}$

If $f^{(n)}(x)$ exists and $|f^{(n)}(x)| \le C$

then c_k decays faster than $\frac{1}{k^n}$

In graph Laplacian case similar result holds.

Prop. Suppose $f = \sum_k c_k \psi_k$ and $f^T L f < \delta$, then $c_k^2 < \frac{\delta}{1-\lambda_k}$ for $k > 1$

$1 = \lambda_1 \ge \lambda_2 \cdots \ge \lambda_n \ge 0$. are eigvalues of $P = D^{-1}W$. $\quad (L = D - W)$

Rmk. $f^T L f = \frac{1}{2} \sum w_{ij} (f_i - f_j)^2$

eg. graph 1-D grid

$$f^T L f \approx -\int_0^1 f(\Delta f)\,dx = \int |\nabla f|^2 dx$$

(consistency)

pf: $L = D - W = D(I-P)$ $P\psi_k = \lambda_k \psi_k$.

Then $f^T L f = f^T D \underbrace{(I-P)}f$

$$(I-P)f = \sum_k c_k (I-P)\psi_k = \sum_k c_k (1-\lambda_k)\psi_k$$

$$\begin{aligned} f^T L f &= (\sum c_k \psi_k)^T D (\sum c_k (1-\lambda_k)\psi_k) \\ &= \sum_{l,k} c_k c_l (1-\lambda_l)\, \psi_k^T D \psi_l \qquad (\Psi^T D \Psi = I) \\ &= \sum_{l,k} c_k c_l (1-\lambda_l)\, \delta_{kl} \\ &= \sum_k c_k^2 (1-\lambda_k) < \delta \qquad c_k^2(1-\lambda_k) < \delta \text{ for all } k. \end{aligned}$$

$\Rightarrow c_k^2(1-\lambda_k)$ decays faster than $\sqrt{\frac{1}{1-\lambda_k}}$

Rmk. Suppose $\lambda_k \downarrow 0$ fast, and δ small
then c_k has a good bound and decays to 0 fast.

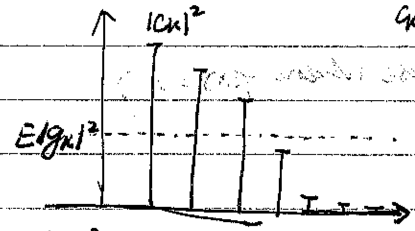
Recall. given $x = f + \varepsilon$ $f, x, \varepsilon \in \mathbb{R}^n$
f is the unknown true signal.
$\varepsilon \sim N(0, \sigma^2 I)$ $\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$

Proposal. $\hat{f} = Px$
$\varepsilon = \sum g_k \psi_k$ $|g_k|$ does not decay
$= \Psi g$ for $g \in \mathbb{R}^n$
if Ψ is unitary. then $g \sim N(0, \sigma^2 I)$

To simplify analysis, as $\Psi^T D \Psi = I$.
we assume $D = k_0 I$ (a regular graph)
and then $I = k_0 \Psi^T \Psi$, i.e. $\sqrt{k_0}\,\Psi$ is unitary

$g = \Psi^{-1}\varepsilon \quad g_k \sim N(0, \nu_0 \sigma^2)$

c_k^2 bdd by $\frac{\delta}{1-\lambda_k}$



$E|g_k|^2 = \nu_0 \sigma^2$

$SNR = \frac{|c_k|^2}{E|g_k|^2}$
(signal noise ratio)

$\cancel{\hat{f} = Pf} = P($

$\hat{f} = Px = P(f+\varepsilon) = P(\sum_k (c_k + g_k)\psi_k)$

$= \sum_k (c_k+g_k)\lambda_k \psi_k$

Bias: $E\hat{f} - f = \cancel{\theta}\ \sum_k c_k \lambda_k \psi_k - \sum c_k \psi_k$

$= -\sum_k c_k(1-\lambda_k)\psi_k.$

Variance: $E\|\hat{f} - E\hat{f}\|^2 = E\|P\varepsilon\|^2$

$= E\|P(\sum_k g_k \psi_k)\|^2$

$= E\|\sum_k g_k \lambda_k \psi_k\|^2$

$= \sum_k \lambda_k^2 \cdot \sigma^2 << \sigma^2 \cdot n$

$\|Bias\|^2 = \sum_k c_k^2 (1-\lambda_k)^2 \cdot \frac{1}{\nu_0} \leq \sum_k c_k^2 (1-\lambda_k) \cdot \frac{1}{\nu_0} \leq \frac{\delta}{\nu_0}$

MLM.

given $\{x_i\}_{i=1}^n \subseteq \mathbb{R}^D$

$\hat{x}_i = \frac{\sum_{j=1}^n w_{ij} \cdot x_j}{\sum_{j=1}^n w_{ij}}$ (denoised patch)

<u>Rmk</u>. $\hat{x}_i(1) = \frac{\sum_{j=1}^n w_{ij} x_j(1)}{\sum_j w_{ij}}$

$f: i \to x_i(1) \qquad \hat{f} = (D^{-1}W) \cdot f$

<spectral filtering>

$P = \Phi \Lambda \Phi^T$

$\hat{f} = \cancel{\sum_k \lambda_k \psi_k} \; Px$

$= \sum \lambda_k \psi_k \langle \varphi_k, x \rangle$ (special case where $f(\lambda) = \lambda$).

generalized version.

$\sum_k f(\lambda_k) \langle \varphi_k, x \rangle \psi_k$.

$SNR_k = \frac{|c_k|^2}{\sigma^2}$

Rmk. Px. global Fourier Transform

The "optimal" denoising strategy. : wavelet shrinkage

$x \rightarrow c_k(x)$ wavelet transform

$x \sim \sum c_k(x) \psi_k$

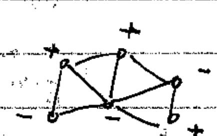
$\rightarrow$ change c_k to be cut off at δ. $T_\delta(c_k) = \hat{c}_k$

$\rightarrow \hat{x} \sim \sum_k \hat{c}_k \psi_k$ $= \begin{cases} c_k & \text{if } |c_k| \geq \delta \\ 0 & \text{else} \end{cases}$

. If f is "smooth", we would expect it to be approximately "flat" in a small interval $\Rightarrow$ we look at f in small windows

Synchronization.

$G = (V, E)$



goal : recover unknown signs on V.

$z_i \in \{-1, 1\}$

from pairwise measurement on E

$(i,j) \in E$. $G_{ij} = z_i z_j = \begin{cases} 1 & z_i = z_j \\ -1 & \text{otherwise} \end{cases}$

A is the adjacency matrix of G.

$G = (z z^T) \odot A \quad = \cancel{z_i z_j \odot A} \; (z_i z_j A_{ij})$

$\downarrow$

$(A \odot B)_{ij} = A_{ij} \cdot B_{ij}$

- We can recover the unknown signs by simply exploring the graph. (when the graph is connected. + there exists "true" signs (without conflict))
- can recover up to flipping all signs in a connected component.

Next. we corrupt some of the observations. $\Rightarrow$ which may result in contradiction

$$G_{ij} = \begin{cases} z_i z_j & (i,j) \in E \quad (A_{ij} = 1) \\ \text{Bernoulli}(\tfrac{1}{2}) & (A_{ij} = 0) \end{cases} \qquad W_{ij} = \begin{cases} 1 & \text{with prob } \tfrac{1}{2} \\ -1 & \text{prob } \tfrac{1}{2} \end{cases}$$

$$G = zz^T \odot A + W \odot (1 \cdot 1^T - A)$$

Task: given G, need to recover z.

Method: take the first eigenvector of G, v_1,

$$\hat{z}_i = \text{sgn}(v_1)$$

rmk. SDP relaxation of the problem:

$$\max \ \langle G, X \rangle$$
$$\text{s.t. } X \succeq 0, \ X_{ii} = 1.$$

$$\langle G, X \rangle = Tr(GX)$$

$$X^* = z \cdot z^T$$

$$Tr(GX^*) = Tr(G \cdot z \cdot z^T) = z^T G z$$

$$X^*_{ii} = z_i^2 = 1.$$

relax the rank constrain $\text{rank}(X) = 1$.

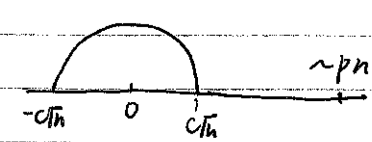
Suppose A is a random graph, $i<j$ i.i.d. (independent of z and W)

$$A_{ij} = \begin{cases} 1 & \text{with prob. } p \\ 0 & \text{with prob. } 1-p \end{cases}$$

$$E_A G = (zz^T) \odot EA + W \odot E(1 \cdot 1^T - A)$$

$$= \underbrace{p \cdot zz^T}_{\text{rank 1 matrix. with } \lambda = pn} + \underbrace{(1-p) W}_{\|(1-p)W\| \lesssim C\sqrt{n}} \qquad C = (1-p)\sqrt{\text{Var}(W_{ij})} \cdot 2$$

when pn is large enough, ($pn > \tilde{C}\sqrt{n}$)
the correlation between the eigenvector of the large eigenvalue ($\sim pn$) and the eigenvector of λ approaches 1.



determinstic $A_{ij}=1$

$$\frac{\#\{\text{good edge } (i,j)\}}{n^2} \sim p$$

eg1. $\mathbb{Z}_2$ signs

eg2. S_1 signs $\theta \in [0, 2\pi]$

$$\theta_{ij} = \begin{cases} t_i - t_j & \text{w.p. } p \\ U([0,2\pi)) & \text{w.p. } 1-p \end{cases}$$

$$z_i = e^{it_i}$$

$$e^{i(t_i - t_j)} = z_i \cdot \bar{z_j}$$