Graph Signal Processed

Saturday, January 21, 2023 6:14 PM

Key ingredients of GSP :-

the mathematics and formulation.

Concept of Shift in DSP :- then developing

a corresponding notion of shift in GSP.

time signals $\longrightarrow$ $\boxed{\text{DSP}}$

Signal samples $\longrightarrow$ $\boxed{\text{GSP}}$

indexed by nodes

of a Graph

At high level DSP and GSP study —

1. Signals and their representation
2. Systems that process signals

   filters.
3. Signal transforms

   Z transform

   Fourier transform
4. Sampling of signals

ects.

$S_n \;,\; n = 0, 1, \cdots, N-1 \,.$

We restrict ourselves $\rightarrow$ N samples.

finite impulse response (FIR.) +

$$S = \{ S_n : n = 0, 1, \cdots, N-1 \}$$ $\leftarrow$ time sample.

$\downarrow$ Z transform

$\Big\downarrow$ Z transform

$S(z)$ $\leftarrow$ ordered set of time samples.

$z^{-1} \leftarrow$ Shift / delay

$$S(z) = \sum_{n=0}^{n-1} S_n z^{-n}$$

$$= S_0 z^{-0} + S_1 z^{-1} + \cdots + S_{n-1} z^{-(n-1)}$$

$$= S_0 + S_1 z^{-1} + \cdots + S_{n-1} z^{-(n-1)}$$

Formal polynomial representation of the signal that is useful in studying how signals are processed by filters.

Given $S(z)$ we can recover the signal.

Discrete FT :-

DFT of the signal s is $\hat{s} = \{\hat{S}_k : k = 0, \dots, N-1\}$

$$\hat{S}_k = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} S_n e^{-j\frac{2\pi}{N}kn}$$

$\frac{2\pi k}{N} = \Omega_k$ $\leftarrow$ discrete frequencies.

$k = 0, 1, \dots, N-1$

$-j\,2\pi k n$ ... $]^{N-1}$

$k = 0, 1, \ldots, N-1$

$$\left\{ x_k[n] = \frac{1}{\sqrt{N}} e^{-j\frac{2\pi k n}{N}} \quad n = 0, 1, \ldots, N-1 \right\}_{k=0}^{N-1}$$

N signals = $\boxed{\text{Spectral components.}}$

Inverse DFT :-

$$S_n = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} \hat{S}_k e^{j\frac{2\pi k n}{N}}$$

$s = 0, 1, \ldots, N-1$

If $H(z)$ is an FIR filter, it can be represented as,

$$h(z) = \sum_{n=0}^{N-1} h_n z^{-n}$$

$$\boxed{S_{out}(z) = h(z) \cdot S_{in}(z)}$$

$S_{out}(z)$ could result in polynomial in z^{-1} of degree greater than $N-1$, ← Need to consider boundary condition.

leveraging the periodicity,

$$S_n = S_{n \bmod N}$$

Shift filter :-

$$h_{shift}(z) = z^{-1}$$

applying it to a signal $s_{in} = (s_0, s_1, \dots, s_{N-1})$ gives an output

$$s_{out} = h_{shift} \cdot s_{in}$$

$$= (s_{N-1}, s_0, s_1, \dots, s_{N-2})$$

Shift Invariance :-

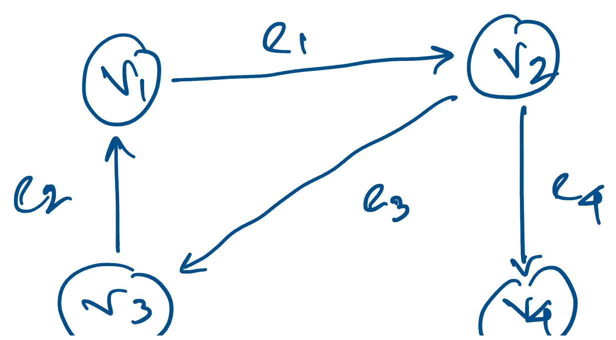
$$z^{-1} \cdot h(z) = h(z) \cdot z^{-1}$$

series combination of the filters are commutative.

$$h * s_{out} = (s_{N-1}, s_0, s_1, \dots, s_{N-2})$$

$$(h_0, h_1, \dots, h_n)$$

∇ = incidence matrix



$$\nabla = \begin{bmatrix} -1 & 1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & -1 & 0 & 1 \end{bmatrix}$$

$$\nabla f(e_{ij}) = f(v_j) - f(v_i)$$

$$\begin{bmatrix} f(2) - f(1) \\ f(1) - f(3) \\ f(3) - f(2) \\ f(4) - f(2) \end{bmatrix}$$

$$\nabla^T \nabla = L \qquad \text{Laplacian}$$

$$\begin{bmatrix} -1 & 1 & 0 & 0 \\ 1 & 0 & -1 & -1 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} -1 & 1 & 0 & 0 \\ 1 & 0 & -1 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & -1 & 0 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} 2 & -1 & -1 & 0 \\ -1 & 3 & -1 & -1 \\ -1 & -1 & 2 & 0 \\ 0 & -1 & 0 & 1 \end{bmatrix}$$

$$= D - A$$

D = degree matrix

A = Adjacency matrix.

Graph Lasso Loss Function Derivation

Friday, March 10, 2023 12:47 PM

The Graph Lasso Loss function is written by

$$\hat{\Theta} = \underset{\Theta}{\operatorname{argmin}} \; tr(S\Theta) - \log|\Theta| + \lambda\|\Theta\|_1$$

One should always demean the data before applying graphical Lasso.

$\boxed{\text{Reason}}$ Graph lasso is all about covariance the concept of which is 'de-meaned'.

Mathematical proof of the Loss function.

Consider the density of the Multivariate Gaussian

$$f(x) = \frac{1}{(2\pi)^{d/2}\,|\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right)$$

If we convert it to likelihood, $\quad z_i \sim N(\mu, \Sigma)$

$$\hat{\Theta} = \underset{\Theta}{\operatorname{argmax}} \; \log f(x_1, \ldots, x_m \mid \Theta)$$

$$= \underset{\Theta}{\operatorname{argmax}} \; \log \prod_{i=1}^{m} f(x_i \mid \Theta)$$

$$= \underset{\Theta}{\operatorname{argmax}} \; \sum_{i=1}^{m} \log f(x_i \mid \Theta)$$

$$\operatorname{argmax} \sum^{m} \log\left[\frac{1}{(2\pi)^{d/2}\,|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x_i-\mu)^T \Theta\,(x_i-\mu)\right)\right]$$

$$= \underset{\Theta}{\operatorname{argmax}} \sum_{i=1}^{n} \log \left[\frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x_i - \mu)^T \Theta (x_i - \mu)\right) \right]$$

$$= \underset{\Theta}{\operatorname{argmax}} \sum_{i=1}^{n} \frac{1}{2} \log |\Sigma^{-1}| + \left[-\frac{1}{2}(x_i - \mu)^T \Theta (x_i - \mu) \right]$$

$$= \underset{\Theta}{\operatorname{argmax}} \frac{n}{2} \log |\Theta| - \frac{1}{2} \sum_{i=1}^{n} (x_i - \mu)^T \Theta (x_i - \mu)$$

$$\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})^T$$ sample cov.

$$= \underset{\Theta}{\operatorname{argmax}} \frac{n}{2} \log |\Theta| - \frac{n}{2} \operatorname{trace}(x^T \Theta x)$$

$$- \frac{n}{2} \operatorname{trace}(x^T x \Theta)$$

$$- \frac{n}{2} \operatorname{trace}(S \Theta)$$

Optimization Essentials

Tuesday, March 14, 2023 2:38 PM

In most ML scenarios, we care about some performance measure P, (test set)

we optimize P indirectly

we reduce a different cost function $J(\theta)$ and hope that optimizing $J(\theta)$ will help us to improve P.

$$J(\theta) = \mathbb{E}_{x,y \sim \hat{p}_{data}} L(f(x;\theta), y)$$

training set

But we are interested to minimize the cost function for the entire population.

like generalization.

But we don't know $P_{data}(x,y)$ we have

$\hat{P}_{data}(x,y) \leftarrow$ samples.

then we replace the true dist by empirical distribution. and minimize the Empirical Risk.

Small batch sizes can offer a regularizing effect

- due to the noise they add to the Learning process.

generalization error is often best for batch size 1.

- Small learning rate is required if the batch size is very small.
  - high variance of the gradient estimate-

Q Why minibatches are sampled randomly?

- for computing unbiased estimates it is important it is required that the samples are indep
- subsequent gradient estimates are also indep from each other.

In the context of NN training,

* We don't care about finding the exact min.
* Only want its value to be reduced sufficiently to obtain a Good generalization error.

<u>SGD</u> :- It is possible to obtain an unbiased estimate of the gradient by taking the average gradient on a minibatch of m examples drawn iid from the data generating dist.

Shifts in GSP

Saturday, January 21, 2023 9:44 PM

Graph Signals :- Signals whose samples are indexed by the nodes of arbitrary graphs.

$$S = (S_0, S_1, \dots, S_{N-1})$$

$$S = [S_0 \; S_1 \; \dots \; S_{N-1}]^T \in \mathbb{C}^N$$

filter h is represented by Matrix H

$$S_{out} = H \cdot S_{in}.$$

RQE

Thursday, January 26, 2023 5:23 PM

→ experiment if possible.
for one example

Not only summary, But
what's new in that domain?

experimental result if it
is not good.

criticize the paper →
- reproducibility →
- experimental details.

presentation :-
very basic + topics in the paper.

$\boxed{\text{Appendix slides}}$

Survey → Break down
on topics.

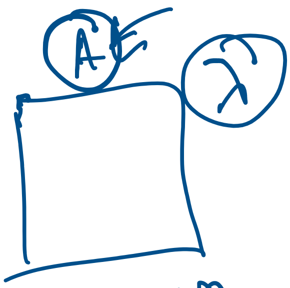
How it is useful?

Clear Basics first. ⇒

check the references

all references.

λ^{-1}



$n \times n$

$Av = \lambda v.$

$\lambda^2 \qquad = 0$

List of Preparations.

Monday, January 30, 2023 7:49 PM

Proof

① In a Symmetric matrix, all the eigen values are real and eigen vectors are also real and form an ortho normal Basis.

② For real symmetric matrix, the AM is equal to geometric multiplicity

③ A symmetric matrix is always diagonalizable.

④ If $\lambda_i \neq \lambda_j$ then

$$s_i \cdot s_j = 0$$

⑤ Why does Laplacian matrix always have a 0 eigen value

⑥ If the characteristic poly $P_A(z)$ and the minimal poly $m_A(z)$ of A are equal then every filter commuting with A is polynomial in A.

$$H = h(A)$$

GSP paper page - 5.

(7) $\text{degree}(h(z)) = \text{degree}(p_A(z)) \leq N-1$

shift invariant filters are polynomials with degree atmost degree $(m_A(z))$.

(7) How to do symmetric normalization of Laplacian?

(8) The roles of time shift in DSP.

all the details of the shift operator.

(9) Proof of convolution.

(10) Discrete FT

(11) IFT

(12) what is FIR.

(13) 6-7 page of GSP paper.

(14) positive semi definite matrix has all eigen values real and non-negative and a full set of orthogonal

vectors can be obtained

$$L = U \Lambda U^T \longrightarrow \text{LU decompo}$$

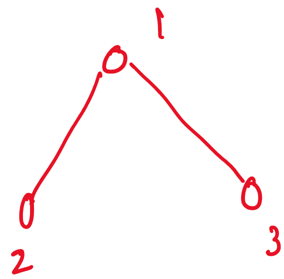
$U \rightarrow$ GFT matrix.

$\boxed{U^T = U^{-1}}$ eigen decom..

SVD

Proof of Laplacian is positive semidefinite

$$L = D - A = BB^T$$



(15) Study about SVD, eigen decom, LU decomposition, Reyleigh quotient.

$$A = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

(16) polynomial approximation with chebyshev's filter.

$$D = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$L = \begin{bmatrix} 2 & -1 & -1 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix}$$

$x^T L x \rightarrow$

$L = U \Lambda U^T$

$$\boxed{U^T L U = \Lambda}$$

$TV(U_k) = \lambda_k$

(17) Why faithfulness assumption is required for obtaining Graph from data ??

Graph Signal Processing Lecture.

Wednesday, February 1, 2023 2:32 PM

1. Graph Frequency
2. Graph Transforms and Filters.
3. Sampling of Graph signals.
4. Learning graphs from data

Data that is on networks.

two types of network
- Physical Network
  - internet
  - power.
- logical network
  - social network
  - The web.
  - . . . . .

How do we take into the structure of these data sets (information encoded in the social network who is connected to whom) to make sense of the information that is available in the network.

Example :-

Relative positions of sensors, temperature

Questions :- Does temperature vary smoothly?

do you see anomalies? or diff behaviours that is not expected.

Social network :- Information about the connection (age, income) connection (age, income)

$\boxed{\text{PyGSP}}$

the graph Laplacian $L = D - A$

Degree - Adjacency

how do we go from this notation to frequency notation?

$$L = D - A = U \Lambda U^t$$

Eigen vectors of $L = U = \{u_k\}_{k=1:n}$

Eigen values of L — Diagonal terms of Λ.

Eigen pair system — (λ_k, u_k)

= Fourier basis of the graph signals.

this is called Graph fourier Transform.

Fourier may not be happy about this. →

But there is an analogy. between the Fourier transform frequencies with the graph eigen values and vectors.

Why GFT?

Because it captures the notion of frequency

Because it captures the notion of frequency

Finding eigen values and eigen vectors.

$$\lambda_k = \min_{u \perp u_1, \dots, u_{k-1}} \frac{u^t L u}{u^t u}$$

solving an optimization which involves the Rayleigh quotient. So, if we have

selected eigenvectors 1 through k-1, we are looking for an eigenvector that is orthogonal to all the prev. ones and minimizes the Rayleigh coefficient.

L = Laplacian ← derived from the graph.

$u_1 = 1$

$\lambda_1 = 0$

$$u^t L u = \sum_{i \sim j} w_{ij}(u_i - u_j)^2$$

u_i, u_j are the values of the eigen vectors at two nodes i and j.

Interpretation:- Basis vectors ordered by increasing variation on the graph.

If $\lambda = 0$ then all the eigenvectors are $\begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}$

If $\lambda = 0$ then all the eigenvectors are $\begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$

$u^t L u \leftarrow$ measures the variation of a signal in a graph.

minimum variation signal is $\mathbb{1}$.

there is no variation, all the signals are same.

Interpret

then when find the next eigen vectors while they are orthogonal to $\mathbb{1}$ ...

we are finding signals that have increased variation while orthogonal to the previous one.

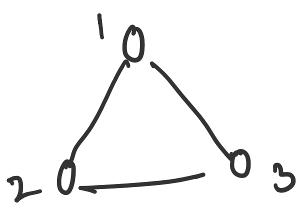
Go through the video $\rightarrow$ 16:00

FB example.

Q open questions about the good definition of frequency ?! (Find those open questions)

low frequency - low variation on the graph.

high frequency - high / rapid variation on the graph



$$L = \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix}$$

$$\begin{vmatrix} 2-\lambda & -1 & -1 \\ -1 & 2-\lambda & -1 \\ -1 & -1 & 2-\lambda \end{vmatrix} = 0 \qquad \lambda = 0, 3, 3$$

Learning Graph.

Sunday, February 26, 2023 1:51 PM

Stationary Graph signal.

A graph signal x is stationary w.r.t the shift S iff $x = Hw$ where $H = \sum_{i=0}^{l-1} h_i S^i$ and w is white.

The covariance matrix C_x of the stationary signal x is a polynomial on S.

$$C_x = \mathbb{E}[x x^T] = \mathbb{E}[Hw (Hw)^T]$$

$$= H\, \mathbb{E}(w w^T) H^T$$

$$= H H^T = H^2$$

$$= \left(\sum_{i=0}^{l-1} h_i S^i\right)^2$$

$C_x S = S C_x$ and eigenvecs(C_x) = eigenvecs(S)

Learning a Graph from Nodal observations.

Given a collection $X := [x_1, \dots, x_p] \in \mathbb{R}^{N \times p}$ of graph signal observations supported on the unknown graph $G(V, \mathcal{E}, W)$ find an optimal S.

$\uparrow$ Shift.

Most classical approaches focus on pairwise similarities

$$\text{sim}(i, j) = f(x_i, x_j)$$

(i) In classical statistics the sim(i,j) is

(i) In classical statistics the sim(i,j) is defined as the correlation coefficient between the nodes.

$$sim(i,j) = \rho_{ij} = \frac{cov[x_i, x_j]}{\sqrt{var[x_i]\, var[x_j]}}$$

then If we really want to know if sim(i,j) = 0 or not, we can do it by performing statistical tests.

$$\hat{\rho}_{ij} = \frac{\hat{c}_{ij}}{\sqrt{\hat{c}_{ii}\, \hat{c}_{jj}}}$$

Edge exist if $0.5 \log\left(\frac{1+\hat{\rho}_{ij}}{1-\hat{\rho}_{ij}}\right) > \frac{Z_{\alpha/2}}{\sqrt{P-3}}$

α = false alarm rate.

The main idea is to obtain the covariance matrix and then sparcify it.

$$y = ax$$

$$\hat{y} = \frac{cov(x,y)}{var(x)} x$$

So, $$y - \hat{y} = ax - \frac{cov(x,y)}{var(x)} x.$$

$$= \left[a - \frac{cov(x,y)}{var(x)}\right] x$$

Idea of the partial correlation :-

For any 2 variables x_1 and x_2, we try to predict x_1 and x_2 with all other variables except x_1 and x_2. then obtain the residual.

If the residual is uncorrelated, or correlated we can say if there is any direct influence from x_1 to x_2 or not.

For partial correlation we look into the precision matrix we try to sparcify the precision matrix.

Learning Graph from Smooth Signals.

Given observations $X = [x_1, \ldots, x_p] \in \mathbb{R}^{N \times p}$, identify a graph G such that signals in X are smooth on G.

Dirichlet energy on the Graph G with Laplacian L

Dirichlet energy on the ...

Search of the shift $S = L$ such that

$$TV(X) = X^T L X \text{ is small.}$$

$$TV(X) = \sum_{p=1}^{P} x_p^T L x_p.$$

$$L^* = \underset{L}{\operatorname{argmin}} \left\{ \sum_{p=1}^{P} x_p^T L x_p + \frac{\beta}{2} \|L\|_F^2 \right\}$$

subject to $\operatorname{trace}(L) = N$??

$$L\mathbf{1} = 0$$

$$L_{ij} = L_{ji} \leq 0 \quad i \neq j$$

Some questions about the GraN DAG paper.

Monday, February 27, 2023 12:46 AM

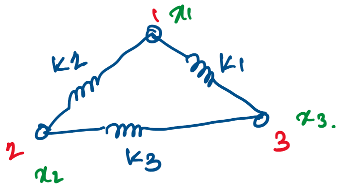
① What different Acyclity constraint they can take account on?

② Why faithfulness assumption is not required here? — Resit

③ to EMRE. Why adding Edges doesn't Reduce the Negative log likelihod?

Spectral Graph Theory (power systems perspective)

Thursday, February 2, 2023 11:54 AM



$$x_3 = \frac{1}{d_3} \sum_{j: i,j \in E} x_j$$

$$= \frac{1}{d_3} (x_1 + x_2)$$

$$= \frac{x_1 + x_2}{d_3.}$$

what is d_i ??

d_i: degree of node (i)

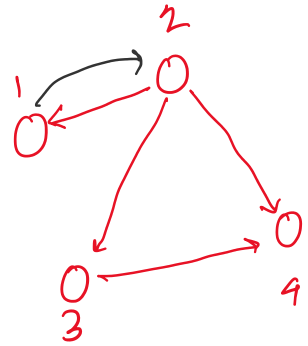
$x_1 = 10$ $x_2 = 10$

$x_3 = 15$

Q Aim it to check, If Graph FT can provide any information about the cascading failure.

$G = (V, E)$ $V = \{v_1, \dots, v_n\}$

$A_G = \{a_{ij}\}$



$K = 1$

	1	2	3	4
1	0	0	0	0
2	1	0	1	1
3	0	0	0	1
4	0	0	0	0

$$A^2 = \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{array}{cccc} 1 & 2 & 3 & 4 \\ \end{array} \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$A^3 =$

$$\begin{bmatrix} 0 & 0 & 0 & 0 \end{bmatrix}$$

$$A^3 = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$tr\left(\sum_{K=1}^{\infty} B^K\right) = 0$$

for acyclic Graphs.

$$tr\,(I-B)^{-1} = d$$

$$tr\,(I-B)^{-1} = tr \sum_{K=0}^{\infty} B^K$$

$$= tr\left[(I) + \sum_{K=1}^{\infty} B^K\right]$$

$$= d + 0$$

$$= d$$

If $eig(B) < 1$ then this results hold true

Is automatically satisfied when B is DAG.

<u>Fault.</u> entries of B^k can easily exceed machine precision for even small values of d.

— lacks numerical stability.

$$tr\,(e^B) = d$$

no cycles $\Rightarrow (B^k)_{ii} = 0 \quad \forall\, k \geq 1$

$$tr(e^B) = tr\left(I + \frac{B^1}{1!} + \frac{B^2}{2!} + \cdots\cdots\right)_\infty$$

$$= d + tr \sum_{i=1}^{\infty} \frac{B^i}{i!} = 0.$$

→ matrix exponential is well-defined for all square matrix.

\# nodes d ↑ → # edges ↑ ⇒ # possible closed walks ↑

$tr(I-B)^{-1}$ → rapidly becomes ill conditioned.

By reweighting the # of lengths-k closed walks by $k!$ this becomes easier to manage.

Famet discrete space

Weighted Adjacency

If B is replaced by W (weighted Adjacency)

this $\boxed{tr\, e^B = d}$ doesn't work

need some alternate criterion.

[However. If W is nonnegative weight matrix this can be proved.]

$$h(W) = tr\left(e^{W \circ W}\right) - d = 0$$

$$\nabla h(W) = e^{(W \circ W)} \odot 2(W)$$

→ counting weighted closed walk. each edge weight is W_{ij}^2.

$h(W) > h(W')$

meaning → (a) W has more cycles than W'

(b) cycles in W are more heavily weighted than in W'.

<u>Equality constrained Program :(ECP)</u>

Plan Tomorrow.

Tuesday, February 28, 2023 12:05 AM

1. Read the causal direction paper for GSP

2. Is there any way to incorporate GSP techniques from causal discovery perspective?

3. Drawbacks in the Gaussian case.

GSP doesn't take the functional Assumption. But it has a smoothness assumption.

Q Is there any relation between smoothness and causality?

4. How to incorporate this optimization procedure

this optimization procedure for cyclic causal discovery?

Constraint $\boxed{B \text{ sparse}}$

and $\boxed{B \text{ needs to be stable}}$

$eig(B) < 1$

can I encode this somehow with the continuous optimization?

then || Gaussian - Non Gaussian problem will be solved.

and || Row permutation problem will also be solved.

But can I obtain the relationship from the logistic regression?

(8)

What does it mean by regression in the case for cyclic SEM. ??