Learning mathematics from S.N Ram

1 optimization

A matrix X is called positive semidefinite if it is symmetric and all its eigenvalues are non-negative. If all eigenvalues are strictly positive then it is called a positive definite matrix

Minimize $C.X(i.e\sum_i\sum_j C_{ij}X_{ij}$) S.T $A_i.X = b_i$ and $X > 0$, If X is symmetric, C is also symmetric and further the optimization form is of SDP(semi definite programming form).Notice that SDP looks remarkably similar to a linear program It is easy to see that a linear program LP is a special instance of an SDP by

$$A_i = \begin{pmatrix} a_{11} & 0 & 0. & . & . \\ 0 & a_{22} & 0. & . & . \\ 0 & 0 & a_{33}. & . & . \\ . & .. & . & . & . \end{pmatrix}$$

$i = 1, 2, 3...$and

$$C_i = \begin{pmatrix} c_{11} & 0 & 0. & . & . \\ 0 & c_{22} & 0. & . & . \\ 0 & 0 & c_{33}. & . & . \\ . & .. & . & . & . \end{pmatrix}$$

$LP = SDP = Min\, C.X\; S.T\; A.X = b_i\; and\; X > 0$ which implies $X = \begin{pmatrix} x_{11} & 0 & 0. & . & . \\ 0 & x_{22} & 0. & . & . \\ 0 & 0 & x_{33}. & . & . \\ . & .. & . & . & . \end{pmatrix}$

1.1 Applications of SDP

In the SDP framework include:linear inequalities, convex quadratic inequalities, lower bounds on matrix norms, lower bounds on determinants of symmetric positive semidefinite matrices, lower bounds on the geometric mean of a nonnegative vector, plus many others. MAX CUT Problem:

1.2 How to solve SDP

There is no finite algorithm for solving SDP. There is a simplex algorithm, but it is not a finite algorithm. There is no direct analog of a "basic feasible solution" for SDP

1.3 Eigen value magics

1.3.1 Nodal voltage using Eigen values

Calculating nodal voltages and branch current flows in a meshed network using Eigen values and eigen vectors of Laplacian matrix (Y_{bus})

$$\begin{gathered} I = YV \\ I = PDP^T V \text{ (since } Y_{bus} \text{ is symmetric matrix)} \\ P^T I = DP^T V \end{gathered}$$

$$\begin{pmatrix} \sum_{k=1}^{N} 1^T I_k \\ \sum_{k=1}^{N} u_{k2} I_k \\ \sum_{k=1}^{N} u_{k3} I_k \\ \cdots \\ .. \end{pmatrix} = \begin{pmatrix} 0 \\ \lambda_2 \sum_{k=1}^{N} u_{k2} V_k \\ \lambda_3 \sum_{k=1}^{N} u_{k3} V_k \\ \cdots \\ .. \end{pmatrix}$$

since, $\lambda_1 = 0$ *and its eigen vector is* 1

$$\begin{pmatrix} \sum_{k=1}^{N} 1^T I_k \\ \sum_{k=1}^{N} u_{k2} I_k \\ \sum_{k=1}^{N} u_{k3} I_k \\ \cdots \\ .. \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & diag(\lambda) \end{pmatrix} \begin{pmatrix} 1^T \\ U \end{pmatrix} V$$

$$V_m = \sum_{k=2}^{N} u_{mk}(b_k/\lambda_k) \text{ as } b_k = \sum_{j=2}^{N} u_{jk} I_j$$

2 Probability

The corner stone of network reduction initiated by Ward [] with the introduction of circuit equivalents using Kron reduction, where mainly focussed on retaining the effect of external circuits on boundary buses

$$Y = \begin{pmatrix} Y_{EE} & Y_{EB} & 0 \\ Y_{BE} & Y_{BB} & Y_{BI} \\ 0 & Y_{IB} & Y_{II} \end{pmatrix} as\ Y_{EB} = Y_{BE};\ Y_{BI} = Y_{IB}$$

$$Y_{red} = Y - 1/Y_{EE} \begin{pmatrix} Y_{EE} \\ Y_{BE} \\ 0 \end{pmatrix} \begin{pmatrix} Y_{EE} \\ Y_{EB} \\ 0 \end{pmatrix}^T = \begin{pmatrix} 0 & 0 & 0 \\ 0 & Y_{BB} - (Y_{BE} * Y_{EB})/Y_{EE} & Y_{BI} \\ 0 & Y_{IB} & Y_{II} \end{pmatrix}$$

$$P_{red} = P - 1/Y_{EE} \begin{pmatrix} Y_{EE} \\ Y_{BE} \\ 0 \end{pmatrix} P_{EE} = \begin{pmatrix} P_{EE} - P_{EE} \\ P_{BB} - Y_{BE} * P_{EE}/Y_{EE} \\ P_{II} \end{pmatrix}$$

However the above method lacks the retainment of system as per the geographical location(zones) and it is more of attention network i.e it equilizes all the buses which are outside of the required zone and the result are dependant on area specific.

New method, called Spectral Clustering proposed for Network reduction based on geographic (nearest neighbor) diatance ,i.e, Clsutering busees which are closes to each other by measuring the distance. Using Gaussian similarity metric, as it gives non linear and normalization function for euclidean distance (0 and 1) $S_{ij} = e^{(-W_{ij}||X_i - X_j||)/\sigma)}$ here σ controls how spread the neighbours are and $Y = D^1 - W$

$X^T SX = \sum_{ij} A_{ij}(X_i - X_j)^2$ where $S = D - A$ Eigen vector corresponding to 2nd least eigen value, gives the bipartie clustering information (two groups), finding K means from K eigen vectors corresponding to K lowest eigen value gives K clusters which formed by gausiian similar matrx.

$$\text{Min } X^T SX$$
$$\text{S.T } X^T X = 1$$

The major drawback of the spectral clustering is, it preserves the local informations, but lacks the inter cluster information (tie line connections).

Further, the development of idea of Network reduction based on PTDF(Power transfer distribution Factor) based comarasion technique.i.e Original network is divided into zones and by minimizing the error between the Zonal PTDF value with reduced network gives the equilant reduced network which arguises the application on online security and market related planning studies.

$P_{Lineflows} = H_{PTDF} P_{Bus_inj}$ and$P_{zonalflow} = T_z H_{PTDF} P_{Bus_inj}$ caused $P_{red} = H_{redPTDF} T_{bz} P_{Bus_inj}$. T_{bZ} matrix be obtained from Zonal information data (Bus aggregation method) and T_z is inter zonal power flow (tie line flow) may be obtained from data(i.e summing the inter zonal power between the aggregated buses) $H_{redPTDF}$ be obtained by minimizing the nter zonal flows of original network to reduced network flows

$$\text{Min } ||P_{red} - P_{zonalflow}||^2$$

Further, $H_{redPTDF}$ can be represented as $Y_{newBR} Y_{new}^{-1}$. As $Y_{newBR} = diag(1/imp_{new}) inc_{new}^T$ and $Y_{new}^{-1} = inc_{new}^{-T} diag(imp_{new}) inc_{new}^{-1}$

Which implies $H_{redptdf} inc_{new} = I$

The basic problem with above method is poorly estimating Y_{new} from $H_{redPTDF}$.

The proposed method slights modify the PTDF reduction method by throwing the laplacian contraints.

3 idea

For a given Network Y_{bus} matrix, using Spectral clustering finding the tie line connected nodes inbetween the clusters. By using Kron reduction(Gauss elimination of nodes) finding the equillant network by removing the clustered nodes leads to the network consists of only Tie line connected nodes and its inter connectivity.

4 notes

Also note that for invertible A,$det(A)$ =det($\mathrm{A}^{-1})^{-1}$ and ,$det(A)$det(A^{-1})=1 If S is invertible, then det(SAS^{-1}) =$det(A)$. Hence similar matrices have the same determinant.Thus it makes sense to define the determinant det T of a linear transformation (i.e., this determinant

4.1 Inverse of the Covariance matrix

$L = Adj - D = \sum^{-1}$ i.e Inverse of variance-covariance matrix of Gausian distribution give Laplace matrix. Where L is semi-definite and 1 is always an eigenvector for the eigenvalue 0 A spanning tree is a subset of Graph G, which has all the vertices covered with minimum possible number of edges. Hence, a spanning tree does not have cycles and it cannot be disconnected.

No of spanning trees $= 1/n(\lambda_1\lambda_2...\lambda_{n-1})$

An eigenvector in general represents a natural mode a system can take. In this case, the system is a graph. And eigenvectors represent these modes the eigenvectors kind of represent the 'frequency' present in a particular weighted graph. They form the x axis if you plot a frequency spectrum. The eigenvalues in turn, represent the 'amounts' of these eigenvectors present in a given graph signal. They form the y axis of the frequency spectrum

Let $X \hookrightarrow \mathcal{N}(\mu, \sigma^2)$ for some $\mu \in \mathcal{R}^d$ and $\sum \in \mathcal{S}^d$. Then , there exists a matrix $B \in \mathcal{R}^{dxd}$ such that if we define $Z = B^{-1}(X - \mu)$, then $Z \hookrightarrow \mathcal{N}(0, 1)$

The central limit theorem states that the distribution of sample means approximates a normal distribution as the sample size gets larger (assuming that all samples are identical in size), regardless of population distribution shape.

Suppose $P_{inj} \sim \mathcal{N}(\mu, \sigma^2)$ and and PTDF matrix is linear matrix which leads to the $P_{line} \sim \mathcal{N}(\mu_{new}, \sigma^2_{new})$ where $\mu_n ew = ptdf * \mu$ and $\sigma_{new} = ptdf * \sigma$

5 Idea of mimicing

5.1 Inverse laplacian

Our goal is to transfer the original space B_{orig} into latent space(reduced) B_{new}. Since Y_{bus} is sparese and inverse Y_{bus} after removing Kth row and Kth column is full matrix.

As we know that $B_{orig} = L$(Laplacian matrix) where $L^{-1} = B^{-1}_{orig} = \sum$(variance co variance of bus nodes), here $\sum$ gives , how two nodes are related.

By using PCA (Prinicple component analysisi) we project the $\sum$ into eigen sapce where theore variance is maximum

$\sum = R_{Bus,NOoffeatures} R^T_{Bus,NOoffeatures}$

By matrix factorization

$R_{Bus,nooffeatures} = P_{bus,N.latentbus} W_{N.latentbus,Nooflatentbus}$

$B_{new} = W^{-T}_{N.latentbus,N.latentbus} W^{-1}_{N.latentbus,N.latentbus} = W^{-2}_{N.latentbus,N.latentbus}$

$B = \phi B_{new} \phi^1$ where $\phi = transformingmatrixfromorigtolowdimenand \phi^1$ is vice versa

$L = D - A$

$L^{-1} = \sum$

$\sum = (D - A)^-1 = D^{-1} + D^{-1}(I - AD^{-1})^{-1} AD^{-1}$

$\sum = D^{-1} P D^{-1}$

AS $P = (I + (I - AD^{-1})^{-1} A)$

$\frac{\partial \sum}{\partial P} = D^{-1} \frac{\partial P}{\partial A} D^{-1} = D^{-1}((I - AD^{-1})^{-1} - A(I - AD^{-1})^{-1} D^{-1} (I - AD^{-1})^{-1}) D^{-1}$

let $D = I$

$\frac{\partial P}{\partial A} = (I - A)^{-1} - A(I - A)^{-1}(I - A)^{-1} = (I - A)^{-1} = P$

$A = log(P)$

Min $||W_1 B_{new}^{-1} W_2 P_{bus} - \theta||^2$
S.T $W_1 ij > 0$ and $W_2 ij > 0$
projection matrix should be positive values only

Further, $W_2 P_{bus} - P_{new}$=0
i.e projection matrix from original p_{bus} to latent space p_{new}

Backpropogation idea envisaged to solve the above optimization problem.

$$Error = (W_1 B_{new}^{-1} W_2 P_{bus} - \theta)$$

$$\frac{\partial L}{\partial W_1} = Error(B_{new}^{-1} W_2 P_{bus})^T$$

$$\frac{\partial L}{\partial W_2} = B_{new}^{-1} {W_1}^T Error P^T$$