EE 771 : Recent Topics in Analytical Signal Processing
Paper Review 1

Arka Sadhu - 140070011

March 21, 2018

Q1

The paper has quite a few differences from the previous literature in the field. In particular:

- Earlier works use some or the other kind of fitness score and compare pairwise nodes and set a weight corresponding to the edge joining the two nodes. The fitness score in some sense evaluates the smoothness of the graph signal.
- The fitness score can either be from a regression model, or correlations between wavelet coefficients or PCA applied to matrices.
- The above class of algorithms only consider the local structure of the graph (pairwise correlation) and not the global structure. Also the paper uses factor analysis which can reveal a simple linear statistical model between the graph signal and the latent variables.
- Another class of algorithms tackle a similar problem of multiple kernel learning. But they use some priors to construct the initial graph and then use smoothness constraint to refine the graph. However, this paper makes no such assumptions and the graph is learnt only through the signal observations.
- The problem of learning graphical models from observed data and inferring graph structure for gaussian markov random fields are similar problems. However, it is known that there is a one-to-one correspondence between the algorithms for solving the above to partial correlations of random variables and hence it again collapses to the case of learning pairwise edge weights.
- Also, the learned graphical model (inverse correlation matrix) cannot be easily projected to get a laplacian which has reuqirements like having row sum as 0, non-positive diagonal entries. Therefore learning the inverse correlation matrix doesn't reveal the global structure of the graph.
- This paper, tries to jointly learn the link between the global smoothness and the graph topology using optimization techniques and this joint learning is what makes the work done novel.

Q2

The smoothness of the graph signal is captured in equation (4) of the paper.

- The paper uses a factor analysis model and uses the following representation:

$$x = \chi h + u_x + \epsilon$$

- Here $x \in \mathbb{R}^n$ is the observed graph signal, $h \in \mathbb{R}^n$ represents the latent variable which controls the graph signal x through the eigen-vector matrix χ. $u_x \in \mathbb{R}^n$ is the mean of x.
- Further it is assumed that the latent variable h has the following distribution:

$$h \sim \mathcal{N}(0, \Lambda^{\dagger})$$

- Here $\Lambda^\dagger$ is the moore-penrose inverse of the eigen-value matrix Λ.
- Clearly, h will have a higher probability of taking a lower value. x is proportional to χh. χ has the usual fourier transform interpretation. Hence smaller values of h implies the graph signal x has more lower frequency components. And thus the smoothness of the graph signal is captured. Also higher the eigen-value, smaller is the standard deviation corresponding to that component and therefore more condensed towards zero. In other words, the higher frequency components have very small contribution.
- We note that any distribution of h which has condensed probability mass near zero is a suitable candidate and can be readily replaced with the gaussian distribution and still capture the smoothness.

Q3

- The input graph model in eq(4) is:

$$x = \chi h + u_x + \epsilon$$

- Here $x \in \mathbb{R}^n$ is the observed graph signal, $h \in \mathbb{R}^n$ represents the latent variable which controls the graph signal x through the eigen-vector matrix χ. $u_x \in \mathbb{R}^n$ is the mean of x. ϵ is the noise that exists in capturing the graph signal.
- $\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2 I_n)$ and $h \sim \mathcal{N}(0, \Lambda^\dagger)$
- The authors claim that the input graph model is generalization of the classical factor analysis model. When representing the graph signal in terms of the eigen-vector matrix χ the model implicitly relates the topology of the graph to the properties of the graph signals. Also it can be interpreted as the fourier basis of the graph signal.
- The distribution of h captures the smoothness of the signal (described in detail in q2).
- Also, in the noise free scenario, x can be seen as a gaussian markov random field with respect to the graph G and the laplacian L being the precision matrix.
- The model also generalizes well in the case of noise.
- In another work, it is shown that signal represneation with the classical factor analysis model provides a probabilistic interpretation of the representation learned by the PCA.
- The distribution of $x|h$ and x are as follows:

$$x|h \sim \mathcal{N}(\chi h + u_x, \sigma_\epsilon^2 I_n)$$

$$x \sim \mathcal{N}(u_x, L^\dagger + \sigma_\epsilon^2 I_n)$$

- When h is given then since ϵ is gaussian distributed, $x|h$ simply becomes a shifted gaussian.
- When h is not given then the x becomes the sum of two gaussians and therefore its distribution is the convolution of the two gauassians and thus the variances add up.

Q4

- The main objective to be optimized is that in eq(15) which is:

$$\min_{L,y} ||x - y||_2^2 + \alpha y^T L y \tag{1}$$

  Here $y = \chi h$

- We need to jointly optimize for both L, y such that y is close to the observed signal x and also that y remains smooth.

- To solve the above 1 the following objective is used:

$$\begin{aligned} \min_{L \in \mathbb{R}^{nxn}, Y \in \mathbb{R}^{nxp}} \quad & ||X - Y||_F^2 + \alpha tr(Y^T L Y) + \beta ||L||_F^2 \\ \text{subject to} \quad & tr(L) = n \\ & L_{ij} = L_{ji} \leq 0, i \neq j, \\ & L \cdot 1 = 0 \end{aligned} \tag{2}$$

- Here $X \in \mathbb{R}^{nxp}$ is the data matrix containing p samples as columns. α and β are two regularization constants. $tr(.)$ is the trace of the matrix. $||.||_F$ is the frobenius norm.
- The first constraint ensures that no trivial solutions for L are found. It also fixes the L1 norm of the laplacian.
- The second and third constraints ensure that the L matrix is always positive semidefinite and also a valid laplacian (since laplacian is given by $D - W$ with all entries of W positive and therefore non-diagonal entries of L are negative or 0.)
- The first and second term in the objective comes directly from the 1
- The last term of frobenius norm of laplacian acts as another regularization to control the distribution of the off-diagonal elements of the laplacian. It is also noted that the when Y is fixed the regularization is akin to that of penalties in elastic net.

Q5

The optimzation problem in 2 is not jointly convex in L and Y. Therefore the authors adopt a scheme of alternating minimization where at each step one variable is fixed and the other variable is solved.

- First they initialize Y as the signal observation X. We note that from the factor analysis method Y can be considered some form of average of the graph signals and therefore the initialization is justified.
- Thus at the first step the following optimization problem is solved for L:

$$\begin{aligned} \min_{L} \quad & \alpha tr(Y^T L Y) + \beta ||L||_F^2 \\ \text{s.t.} \quad & tr(L) = n \\ & L_{ij} = L_{ji} \leq 0, i \neq j \\ & L \cdot 1 = 0 \end{aligned} \tag{3}$$

- In the second step, L is fixed to the value from above and the following optimization problem is solved for Y:

$$\min_{Y} ||X - Y||_F^2 + \alpha tr(Y^T L Y) \tag{4}$$

- It is further noted that both the problems in 3 and 4 can be casted as convex optimization problems with unique minimizers.