Lagrangian relaxation for tree-structured covariance estimation

Xiyang Dai, Zebao Gao and Hao Zhou
Dept. of Computer Science

1. Tree-structured Covariance Estimation

Understanding similarities in expression profiles from a set of samples, e.g., multiple brain regions, is valuable to help understand many biological processes. A natural way to analyze this kind of similarities is to model a tree-structured covariance estimation. However, this kind of model turns out to be an non-convex optimization problems shown in recent paper from Hector etc [1]. In their paper, they have proposed a mixed-integer programming (MIP) approach to solve this non-convex optimization problem. Specifically, they estimate a covariance matrix from observations of p continuous random variables encoding a stochastic process over a tree with p leaves. To formulate the estimation problem as instances of weel-studied numerical optimization problems, they used linear combinations of rank-one matrices indicating object partitions. Although they have shown great performance using MIP, this non-convex optimization problem for tree-structure covariance estimation is still far from solved.

2. Lagrangian Relaxation

Recently, in Natural Language Processing (NLP) filed, Lagrangian relaxation methods are employed for solving non-convex optimization problems .

Recently, there is an increasing trend in applying Lagrangian relaxation methods to solve non-convex optimization problems, especially in NLP field. It has been successfully applied to several NLP inference problems such as Part-of-speech tagging [2]. But more studies are still desirable to conduct to evaluate its practicability in the tree-structured covariance estimation in our Bioinformatics problems.

3. Our Project

In this project, we will try to see whether Lagrangian relaxation method can be applied to solve this kind of non-convex optimization problem, and try to apply this tree-structured estimation method to Bio-related topics. The main challenges of our project include:

1 understanding the existing tree-structured covariance estimation problem;
2 understanding the Lagrangian relaxation method;
3 try to apply the method from NLP field to bio-related topics;
4 implementation of these methods.

4. Formulation of Problem

A set of nested partitions of objects can be defined by a rooted tree such that each node in the tree corresponds to a subset of these objects. The following definitions and the theorem introduces the most important terms and their relationship.

Definition 1 (Tree-structured Covariance Matrix)**.** *A tree-structured covariance matrix B is a matrix such that each entry B_{ij} is the sum of branch lengths for the path starting at the root and ending at the last common ancestor of leaves i and j.*

Definition 2. *Partition Property A basis matrix V of size p-by-(2p-1) with entries in 0, 1 and unique columns has the partition property for trees of size p if it satisfies the following conditions:*

1) *V contains the vector of all onews $e = (1, 1, ..., 1)^T \in R^p$ as a column; and*
2) *for every column w in V with more than one non-zero entry, it contains exactly two columns u and v such that $u + v = w$.*

Theorem 1 (Tree Covariance Representation)**.** *A matrix B is a tree-structured covariance matrix if and only if $B = VDV^T$ where D is a diagonal matrix with nonnegative entries and the basis matrix V has the partition property.*

Now, we formulate our problem as follows: Give a sample covariance matrix S, we find the nearest tree-structured covariance matrix in Forbenius norm $|| \bullet ||_F$. Let s be the vextorization of symmetric matrix S such that $||S||_F = ||s||_F$, then the problem we are solving is to find the solution to the following objective function:

$$\min_{b \in R^{p(p+1)/2}, \rho \in R^{\overline{p}}} f(b) = \frac{1}{2} b^T b - s_T b$$

subject to the following constraints:

$$
\begin{aligned}
B_{ij} &\geq 0 \ \forall \ i, j && (1)\\
B_{ii} &\geq B_{ij} \ \forall \ i \neq j && (2)\\
B_{ij} &\geq B_{ik} - (1 - \rho_{ijk1})M && (3)\\
B_{ik} &\geq B_{jk} - (1 - \rho_{ijk1})M && (4)\\
B_{jk} &\geq B_{ik} - (1 - \rho_{ijk1})M && (5)\\
B_{ik} &\geq B_{ij} - (1 - \rho_{ijk2})M && (6)\\
B_{ij} &\geq B_{jk} - (1 - \rho_{ijk2})M && (7)\\
B_{jk} &\geq B_{ij} - (1 - \rho_{ijk2})M && (8)\\
B_{jk} &\geq B_{ij} - (\rho_{ijk1} + \rho_{ijk2})M && (9)\\
B_{ij} &\geq B_{ik} - (\rho_{ijk1} + \rho_{ijk2})M && (10)\\
B_{ik} &\geq B_{ij} - (\rho_{ijk1} + \rho_{ijk2})M && (11)\\
\rho_{ijk1} + \rho_{ijk2} &\leq 0 && (12)\\
\rho_{ijk1}, \rho_{ijk2} &\in \{0, 1\} \ \forall \ i > j > k && (13)
\end{aligned}
$$

5. Our Approach

The problem in the previous section can be solved using Mixed-Integer Programming (MIP).

In this project, we solve the problem using Lagrangian Relaxation.
Then our objective function becomes: $\min_{b \in R^{p(p+1)/2}, \rho \in R^{\overline{p}}} f(b) = \frac{1}{2} b^T b - s_T b + \sum_{i,j,k} constraints(....)$

Further more, we can introduce a new $\rho ijk3$ where $\rho_{ijk1} + \rho_{ijk2} + \rho_{ijk3} = 1$ and $\rho_{ijk1}, \rho_{ijk2}, \rho_{ijk3} \in [0, 1]$. To make the solution of ρ_{ijk1}, ρ_{ijk2} and ρ_{ijk3} tend to be one of the value in the set 0, 1, we construct the following two functions:

1) $g_1(\rho) = \sum_{\rho} M\rho(1-\rho)$
2) $g_2(\rho) = -\sum_{\rho} M\rho log(\rho)$

Then we update the constraints as follows:

$$B_{ij} \geq 0 \ \forall \ i, j \tag{14}$$
$$B_{ii} \geq B_{ij} \ \forall \ i \neq j \tag{15}$$
$$B_{ij} \geq B_{ik} - (1 - \rho_{ijk1})M \tag{16}$$
$$B_{ik} \geq B_{jk} - (1 - \rho_{ijk1})M \tag{17}$$
$$B_{jk} \geq B_{ik} - (1 - \rho_{ijk1})M \tag{18}$$
$$B_{ik} \geq B_{ij} - (1 - \rho_{ijk2})M \tag{19}$$
$$B_{ij} \geq B_{jk} - (1 - \rho_{ijk2})M \tag{20}$$
$$B_{jk} \geq B_{ij} - (1 - \rho_{ijk2})M \tag{21}$$
$$B_{jk} \geq B_{ij} - (1 - \rho_{ijk3})M \tag{22}$$
$$B_{ij} \geq B_{ik} - (1 - \rho_{ijk3})M \tag{23}$$
$$B_{ik} \geq B_{ij} - (1 - \rho_{ijk3})M \tag{24}$$
$$\rho_{ijk1} + \rho_{ijk2} + \rho_{ijk3} = 1 \tag{25}$$
$$\rho_{ijk1}, \rho_{ijk2}, \rho_{ijk3} \in [0, 1] \ \forall \ i > j > k \tag{26}$$

And the object function can be:

1) $\min_{b \in R^{p(p+1)/2}, \rho \in R^{\overline{p}}} f(b) = \frac{1}{2} b^T b - s_T b \ + g_1(\rho)$
2) $\min_{b \in R^{p(p+1)/2}, \rho \in R^{\overline{p}}} f(b) = \frac{1}{2} b^T b - s_T b \ + g_2(\rho)$

References

[1] Héctor Corrada Bravo, Stephen J. Wright, Kevin H. Eng, Sunduz Keles, and Grace Wahba. Estimating tree-structured covariance matrices via mixed-integer programming. In *AISTATS*, JMLR Proceedings, pages 41–48, 2009.

[2] Alexander M. Rush and Michael Collins. A tutorial on dual decomposition and lagrangian relaxation for inference in natural language processing. *Journal of Artificial Intelligence Research*, 2012.