Online Companion to the Paper 'Transformation of Acyclic Phase Type Distributions for Correlation Fitting'

Peter Buchholz, Iryna Felko, and Jan Kriege

Department of Computer Science, TU Dortmund
{peter.buchholz,iryna.felko,jan.kriege}@udo.edu

Abstract. This note includes some additional details as an extension to the paper "Transformation of Acyclic Phase Type Distributions for Correlation Fitting" [1].

1 Introduction

This online companion adds information and details that could not be included in the paper due to space restrictions. In this respect it is only fully understandable with the knowledge of the original paper [1]. In Sect. 2 we prove that the hyperexponential representation is the representation that allows for a maximal first joint moment when expanding the Phase-type distribution into a MAP. Sect. 3 deals with the expansion of the state space of PH distributions. In particular, we show that the state space expansion proposed in [1] results in a representation with a first joint moment which is not smaller than the one of the original PH distribution.

2 The General Case

In [1] we defined the following transformation which modifies the representation but not the distribution and still results in an APH. The following equations apply the transformation to two states $i < j$ and we assume that for $i < j$ $\lambda_i \leq \lambda_j$. $\mathbf{p}^\delta$ and $\mathbf{D}_0^\delta$ are the vector and matrix after the transformation with parameter δ has been applied.

$$\mathbf{p}^\delta(k) = \begin{cases} \mathbf{p}(i) + \delta & \text{for } k = i \\ \mathbf{p}(j) - \delta & \text{for } k = j \\ \mathbf{p}(k) & \text{otherwise} \end{cases} \qquad \lambda_{k,l} = \begin{cases} \lambda_{i,j}\frac{\mathbf{p}(i)}{\mathbf{p}(i)+\delta} - \frac{(\lambda_j-\lambda_i)\delta}{\mathbf{p}(i)+\delta} & \text{for } k = i \text{ and } l = j \\ \lambda_{i,l}\frac{\mathbf{p}(i)}{\mathbf{p}(i)+\delta} & \text{for } k = i \text{ and } l > i \wedge l < j \\ \lambda_{i,l}\frac{\mathbf{p}(i)}{\mathbf{p}(i)+\delta} + \lambda_{j,l}\frac{\delta}{\mathbf{p}(i)+\delta} & \text{for } k = i \text{ and } l > j \\ \lambda_{k,i}\frac{\mathbf{p}(i)+\delta}{\mathbf{p}(i)} & \text{for } k < i \text{ and } l = i \\ \lambda_{k,j} - \lambda_{k,i}\frac{\delta}{\mathbf{p}(i)} & \text{for } k < i \text{ and } l = j \\ \lambda_{k,l} & \text{otherwise} \end{cases} \tag{1}$$

The diagonal entries λ_k are not modified by the transformation. For the exit vector $\mathbf{d}_1$, $\mathbf{d}_1^\delta(k) = \mathbf{d}_1(k)$ for $k \neq i$ and

$$\mathbf{d}_1^\delta(i) = \frac{\mathbf{p}(i)\mathbf{d}_1(i) + \delta\mathbf{d}_1(j)}{\mathbf{p}(i)+\delta} = \mathbf{d}_1(i) + \delta\frac{\mathbf{d}_1(j) - \mathbf{d}_1(i)}{\mathbf{p}(i)+\delta}. \tag{2}$$

To compute a valid APH, parameter δ has to be chosen from the following interval to assure that the rates and probabilities remain non-negative.

$$\left[\max\left(-\mathbf{p}(i), \min_{l>j,\lambda_{j,l}>0}\left(-\frac{\mathbf{p}(i)\lambda_{i,l}}{\lambda_{j,l}}\right), -\frac{\mathbf{p}(i)\mathbf{d}_1(i)}{\mathbf{d}_1(j)}\right),\ \min\left(\mathbf{p}(j), \min_{k<i,\lambda_{k,i}>0}\left(\frac{\mathbf{p}(i)\lambda_{k,j}}{\lambda_{k,i}}\right), \frac{\mathbf{p}(i)\lambda_{i,j}}{\lambda_j - \lambda_i}\right)\right] \quad (3)$$

If $\mathbf{d}_1(j) = 0$, then the last term for the lower bound becomes $-\infty$ and is not used. If $\lambda_j = \lambda_i$, then the last term in the upper bound evaluates to ∞. Furthermore, if δ is set to $-\mathbf{p}(i)$, then the rates of the new APH become infinite.

Using this transformation we proposed the following theorem that will be proved in the remainder of this section.

Theorem 1. *If an APH can be transformed into an hyperexponential representation with* $\mathbf{p}(i) > 0$ *using similarity transformations (1)-(2), then this representation results in the maximal value* $\mu^*_{1,1}$.

Proof. We first consider $\mu_{1,1}$ for the hyperexponential representation. Let $\mathbf{\Lambda}$ be the vector of phase rates, $\mathbf{\Lambda}^{-1}$ a column vector with λ_i^{-1} in position i and $\mathbf{p}$ the initial vector of the hyperexponential representation. Then $\mathbf{D}_0 = diag(-\mathbf{\Lambda})$, $\mathbf{d}_1 = \mathbf{\Lambda}$ and $\mathbf{M} = (-\mathbf{D}_0)^{-1} = diag(\mathbf{\Lambda}^{-1})$. Let $\mathbf{B}$ be a stochastic matrix such that $\mathbf{D}_1 = diag(\mathbf{d}_1)\mathbf{B} = diag(\mathbf{\Lambda})\mathbf{B}$, $\mathbf{pB} = \mathbf{p}$ and $\mu_{1,1}$ is maximal. $\mu_{1,1}$ can be represented as

$$\mu_{1,1} = \mathbf{p}\mathbf{M}^2\mathbf{D}_1\mathbf{M}\mathbb{I} = \mathbf{p}diag(\mathbf{\Lambda}^{-1})\mathbf{B}\mathbf{\Lambda}^{-1}.$$

Matrix $\mathbf{B}$ that maximizes $\mu_{1,1}$ can be computed from the linear optimization problem with variables $\mathbf{B}(i, j)$ $(i, j = 1, \ldots, n)$, $2n$ linear constraints

$$\mathbf{p}(i) = \sum_{j=1}^{n} \mathbf{p}(j)\mathbf{B}(j, i), \sum_{j=1}^{n} \mathbf{B}(i, j) = 1 \text{ and goal function } \sum_{i=1}^{n} \frac{\mathbf{p}(i)}{\lambda_i} \sum_{j=1}^{n} \frac{\mathbf{B}(i, j)}{\lambda_j}.$$

The maximum is achieved for $\mathbf{B} = \mathbf{I}$ in this case.

Now consider some APH with representation $(\tilde{\mathbf{p}}, \tilde{\mathbf{D}}_0, \tilde{\mathbf{d}}_1)$ which results from the hyperexponential distribution $(\mathbf{p}, \mathbf{D}_0, \mathbf{d}_1)$ using the similarity transformation defined in (1)-(2). These transformations can be collected in a non-singular transformation matrix $\mathbf{C}$ with $\mathbf{C}\mathbb{I} = \mathbb{I}$ such that

$$\tilde{\mathbf{p}} = \mathbf{pC},\ \tilde{\mathbf{D}}_0 = \mathbf{C}^{-1}\mathbf{D}_0\mathbf{C} \text{ and } \tilde{\mathbf{d}}_1 = \mathbf{C}^{-1}\mathbf{d}_1$$

where $\mathbf{D}_0 = diag(-\mathbf{\Lambda})$ and $\mathbf{d}_1 = \mathbf{\Lambda}$. Furthermore, $\tilde{\mathbf{M}} = (-\tilde{\mathbf{D}}_0)^{-1} =$
$\mathbf{C}^{-1}(-\mathbf{D}_0)^{-1}\mathbf{C} = \mathbf{C}^{-1}\mathbf{M}\mathbf{C} = \mathbf{C}^{-1}diag(\mathbf{\Lambda}^{-1})\mathbf{C}$. Now consider the first joint moment of the transformed representation which equals

$$\begin{aligned}\tilde{\mu}_{1,1} &= \tilde{\mathbf{p}}\tilde{\mathbf{M}}^2 diag(\tilde{\mathbf{d}}_1)\mathbf{B}\tilde{\mathbf{M}}\mathbb{I} \\ &= \mathbf{pCC}^{-1}\mathbf{MCC}^{-1}\mathbf{MC}diag(\mathbf{C}^{-1}\mathbf{d}_1)\mathbf{BC}^{-1}\mathbf{MC}\mathbb{I} \\ &= \mathbf{p}diag(\mathbf{\Lambda}^{-1})^2 \underbrace{\mathbf{C}diag(\mathbf{C}^{-1}\mathbf{\Lambda})\mathbf{BC}^{-1}}_{\mathbf{G}} \mathbf{\Lambda}^{-1}.\end{aligned}$$

We have

$$\mathbf{G}\mathbb{I} = \mathbf{C}diag(\mathbf{C}^{-1}\mathbf{\Lambda})\mathbf{B}\mathbf{C}^{-1}\mathbb{I} = \mathbf{C}diag(\mathbf{C}^{-1}\mathbf{\Lambda})\mathbb{I} = \mathbf{\Lambda}$$

and

$$\mathbf{p}\tilde{\mathbf{M}}\tilde{\mathbf{D}}_1 = \mathbf{p}diag(\mathbf{\Lambda}^{-1})\mathbf{G} = \mathbf{p}.$$

Again we can formulate a linear program. However, we cannot assume that matrix $\mathbf{G}$ is non-negative. To obtain a linear program in augmented form we introduce variables $\mathbf{G}^-(i, j)$ which represent the negative part and $\mathbf{G}^+(i, j)$ which represent the positive part. Then we assume $\mathbf{G}^-(i, j), \mathbf{G}^+(i, j) \geq 0$ and define linear constraints for $i = 1, \ldots, n$

$$\sum_{j=1}^{n} (\mathbf{G}^+(i, j) - \mathbf{G}^-(i, j)) = \lambda_i \text{ and } \sum_{j=1}^{n} \frac{\mathbf{p}(j)}{\lambda_j}(\mathbf{G}^+(j, i) - \mathbf{G}^-(j, i)) = \mathbf{p}(i)$$

and the goal function

$$\sum_{i=1}^{n} \frac{\mathbf{p}(i)}{\lambda_i^2} \sum_{j=1}^{n} \frac{(\mathbf{G}^+(i, j) - \mathbf{G}^-(i, j))}{\lambda_j}.$$

Since this is a linear programming problem, the solution is an extremal point of the feasible region [2]. The extremal points are those points where for each i exactly one value $\mathbf{G}^+(i, j)$ is equal to λ_i (or $\mathbf{G}^-(i, j)$ is equal to $-\lambda_i$ which does not result in a valid APH). To maximize the goal function, we have to choose $\mathbf{G}^+(i, i) = \lambda_i$ such that $\mathbf{G} = diag(\mathbf{\Lambda})$ and $\mu_{1,1}$ equals the value for the hyperexponential distribution.

3 State Space Expansion

For cases where a desired autocorrelation cannot be achieved for a given APH distribution we proposed a method to enlarge the state space by adding additional phases to increase the flexibility in [1].

Let $(\mathbf{p}, \mathbf{D}_0)$ be the original n-dimensional APH. We define a $n+1$ dimensional APH $(\mathbf{p}', \mathbf{D}_0')$ with

$$\mathbf{p}'(i) = \begin{cases} \mathbf{p}(i) & \text{if } i < n \\ 0 & \text{if } i = n \\ \mathbf{p}(n) & \text{if } i = n+1 \end{cases} \quad \text{and } \mathbf{D}_0'(i, j) = \begin{cases} \mathbf{D}_0(i, j) & \text{if } i < n \text{ and } j < n \\ \mathbf{a}(i)\mathbf{D}_0(i, n) & \text{if } i < n \text{ and } j = n \\ (1 - \mathbf{a}(i))\mathbf{D}_0(i, n) & \text{if } i < n \text{ and } j = n+1 \\ \mathbf{D}_0(n, n) & \text{if } i, j \in \{n, n+1\} \text{ and } i = j \\ 0 & \text{if } i = n \text{ and } j = n+1 \end{cases} \tag{4}$$

for some vector $\mathbf{a}$ of length n with elements out of $[0, 1]$. It is easy to show that the required relation between the two APHs holds. Although the above transformation works for all vectors $\mathbf{a}$ with elements from $[0, 1]$ we assume in the sequel that $\mathbf{a} = \mathbb{I}$.

Let $(\mathbf{D}_0, \mathbf{D}_1)$ be a MAP expanded from $(\mathbf{p}, \mathbf{D}_0)$ with a maximal first joined moments $\mu^*_{1,1}$. We define a MAP $(\mathbf{D}'_0, \mathbf{D}'_1)$ with matrix $\mathbf{D}'_0$ as in (4) with vector $\mathbf{a} = \mathbb{I}$ and

$$\mathbf{D}'_1(i,j) = \begin{cases} \mathbf{D}_1(i,j) + \frac{\mathbf{D}_1(n,j)}{\sum_{k=1}^{n-1} \mathbf{D}_1(n,k)} \mathbf{D}_1(i,n) & \text{if } i < n \text{ and } j < n \\ 0 & \text{if } i < n \text{ and } j \geq n \\ \frac{\sum_{k=1}^{n} \mathbf{D}_1(n,k)}{\sum_{l=1}^{n-1} \mathbf{D}_1(n,l)} \mathbf{D}_1(n,j) & \text{if } i = n \text{ and } j < n \\ 0 & \text{if } i = j = n \\ 0 & \text{if } i \leq n \text{ and } j = n+1 \\ \lambda_n & \text{if } i = j = n+1 \end{cases}$$

In the following we will show that $(\mathbf{D}'_0, \mathbf{D}'_1)$ describes a valid MAP and that $\mu^*_{1,1} \leq \mu'_{1,1}$, i.e. the transformation results in a representation with a first joint moment which is not smaller than the first joint moment of the original APH.

Let $\mathbf{M}' = (-\mathbf{D}'_0)^{-1}$. We show that $\mathbf{p}'\mathbf{M}'\mathbf{D}'_1 = \mathbf{p}'$ such that $\mathbf{D}'_1$ is a valid matrix for a MAP expanded from APH $(\mathbf{p}', \mathbf{D}'_0)$ and that $\mu^*_{1,1} = \mathbf{p}\mathbf{M}^2\mathbf{D}_1\mathbf{M}\mathbb{I} \leq \mu'_{1,1} = \mathbf{p}'(\mathbf{M}')^2\mathbf{D}'_1\mathbf{M}'\mathbb{I}$. We use the notations $\mathbf{m} = \mathbf{M}\mathbb{I}$, $\mathbf{m}' = \mathbf{M}'\mathbb{I}$ and $\mathbf{n} = \mathbf{p}\mathbf{M}^2$, $\mathbf{n}' = \mathbf{p}'(\mathbf{M}')^2$. The following relations can be shown by simple calculations.

$$\mathbf{m}'(i) = \begin{cases} \mathbf{m}(i) & \text{if } i \leq n \\ \mathbf{m}(n) & \text{if } i = n+1 \end{cases} \quad \text{and } \mathbf{n}'(i) = \begin{cases} \mathbf{n}(i) & \text{if } i < n \\ \mathbf{n}(n) - \frac{\mathbf{p}(n)}{\lambda_n^2} & \text{if } i = n \\ \frac{\mathbf{p}(n)}{\lambda_n^2} & \text{if } i = n+1 \end{cases}$$

This implies

$$\begin{aligned} \mu^*_{1,1} &= \sum_{i=1}^{n} \mathbf{n}(i) \sum_{j=1}^{n} \mathbf{D}_1(i,j)\mathbf{m}(j) \\ &= \sum_{i=n}^{n-1} \mathbf{n}(i) \sum_{j=1}^{n-1} \mathbf{D}_1(i,j)\mathbf{m}(j) + \mathbf{n}(n) \sum_{j=1}^{n} \mathbf{D}_1(n,j)\mathbf{m}(j) + \sum_{i=1}^{n-1} \mathbf{n}(i)\mathbf{D}_1(i,n)\mathbf{m}(n) \\ &\leq \sum_{i=n}^{n-1} \mathbf{n}(i) \sum_{j=1}^{n-1} \mathbf{D}_1(i,j)\mathbf{m}(j) + \left(\mathbf{n}(n) - \frac{\mathbf{p}(n)}{\lambda_n^2}\right) \sum_{j=1}^{n} \frac{\sum_{k=1}^{n} \mathbf{D}_1(n,k)}{\sum_{l=1}^{n-1} \mathbf{D}_1(n,l)} \mathbf{D}_1(n,j)\mathbf{m}(j) + \frac{\mathbf{p}(n)}{\lambda_n^2} \\ &\qquad + \sum_{i=1}^{n-1} \mathbf{n}(i)\mathbf{D}_1(i,n) \sum_{j=1}^{n-1} \frac{\mathbf{D}_1(n,j)}{\sum_{k=1}^{n-1} \mathbf{D}_1(n,k)} \mathbf{m}(j) \\ &= \sum_{i=1}^{n+1} \mathbf{n}'(i) \sum_{j=1}^{n+1} \mathbf{D}'_1(i,j)\mathbf{m}'(j) = \mu'_{1,1} \end{aligned}$$

The transformation from the second to the third line holds because $\sum_{j=1}^{n-1} \mathbf{D}_1(n,j)\mathbf{m}(j) = 1$ such that

$$\mathbf{n}(n) \sum_{j=1}^{n-1} \mathbf{D}_1(n,j)\mathbf{m}(j) = \mathbf{n}(n) \leq \left(\mathbf{n}(n) - \frac{\mathbf{p}(n)}{\lambda_n^2}\right) \sum_{j=1}^{n-1} \frac{\sum_{k=1}^{n} \mathbf{D}_1(n,k)}{\sum_{l=1}^{n-1} \mathbf{D}_1(n,l)} \mathbf{D}_1(n,j)\mathbf{m}(j) + \frac{\mathbf{p}(n)}{\lambda_n^2}$$

and $\mathbf{m}(j) \geq \mathbf{m}(n)$ since $\lambda_n \geq \lambda_j$ for $j < n$) such that

$$\sum_{j=1}^{n-1} \frac{\mathbf{D}_1(n, j)}{\sum\limits_{k=1}^{n-1} \mathbf{D}_1(n, k)} \mathbf{m}(j) \geq \mathbf{m}(n) \sum_{j=1}^{n-1} \frac{\mathbf{D}_1(n, j)}{\sum\limits_{k=1}^{n-1} \mathbf{D}_1(n, k)} = \mathbf{m}(n).$$

To show that $\mathbf{p}'\mathbf{M}'\mathbf{D}_1' = \mathbf{p}'$ let $\mathbf{r} = \mathbf{pM}$ and $\mathbf{r}' = \mathbf{p}'\mathbf{M}$ which implies

$$\mathbf{r}'(i) = \begin{cases} \mathbf{r}(i) & \text{if } i < n \\ \mathbf{r}(n) - \frac{\mathbf{p}(n)}{\lambda_n} & \text{if } i = n \\ \frac{\mathbf{p}(n)}{\lambda_n} & \text{if } i = n+1 \end{cases}$$

Since $\mathbf{p} = \mathbf{r}\mathbf{D}_1$ and $\mathbf{p}' = \mathbf{r}'\mathbf{D}_1'$ we have for $i < n$

$$\begin{aligned}
\mathbf{p}'(i) &= \sum_{j=1}^{n} \mathbf{r}'(j)\mathbf{D}_1'(j,i) \\
&= \sum_{j=1}^{n-1} \mathbf{r}(j)\left(\mathbf{D}_1(i,j) + \frac{\mathbf{D}_1(n,i)}{\sum_{k=1}^{n-1}\mathbf{D}_1(n,k)}\mathbf{D}_1(j,n)\right) + \left(\mathbf{r}(n) - \frac{\mathbf{p}(n)}{\lambda_n}\right) \frac{\sum_{k=1}^{n}\mathbf{D}_1(n,k)}{\sum_{l=1}^{n-1}\mathbf{D}_1(n,l)}\mathbf{D}_1(n,i) \\
&= \sum_{j=1}^{n-1} \mathbf{r}(j)\mathbf{D}_1(j,i) + \frac{\mathbf{D}_1(n,i)}{\sum_{l=1}^{n-1}\mathbf{D}_1(n,l)}\left(\sum_{j=1}^{n-1}\mathbf{r}(j)\mathbf{D}_1(j,n) + \left(\mathbf{r}(n) - \frac{\mathbf{p}(n)}{\lambda_n}\right)\sum_{k=1}^{n}\mathbf{D}_1(n,k)\right) \\
&= \sum_{j=1}^{n-1} \mathbf{r}(j)\mathbf{D}_1(j,i) + \frac{\mathbf{D}_1(n,i)}{\sum_{l=1}^{n-1}\mathbf{D}_1(n,l)}\mathbf{r}(n)\sum_{k=1}^{n-1}\mathbf{D}_1(n,k) = \mathbf{p}(i).
\end{aligned}$$

The transformation form the second last to the last row holds since $\sum_{j=1}^{n} \mathbf{D}_1(n,i) = \lambda_n$ and $\sum_{j=1}^{n-1} \mathbf{r}(i)\mathbf{D}_1(i,n) = \mathbf{p}(n) - \mathbf{r}(n)\mathbf{D}_1(n,n)$. $\mathbf{p}'(n) = 0$ because column n of $\mathbf{D}_1' = \mathbf{0}$ and $\mathbf{p}'(n+1) = \mathbf{r}'(n+1)\mathbf{D}_1'(n+1, n+1) = \mathbf{p}(n)$.

References

1. P. Buchholz, I. Felko, and J. Kriege. Transformation of acyclic phase type distributions for correlation fitting, 2013. Submitted for Twentieth International Conference on Analytical & Stochastic Modelling Techniques & Applications (ASMTA).
2. K. Neumann and M. Morlock. *Operations Research*. Hanser, 2002.