Irrelevant Variability Normalization Based HMM Training Using VTS Approximation of an Explicit Model of Environmental Distortions

Yu Hu1,2 and Qiang Huo1

1Department of Computer Science, The University of Hong Kong, Hong Kong, China
2Department of Electronic Engineering & Information Science, University of Science and Technology of China, Hefei, China
(Emails: jadefox@ustc.edu, quho@cs.hku.hk)

Abstract
In a traditional HMM compensation approach to robust speech recognition that uses Vector Taylor Series (VTS) approximation of an explicit model of environmental distortions, the set of generic HMMs are typically trained from “clean” speech only. In this paper, we present a maximum likelihood approach to training generic HMMs from both “clean” and “corrupted” speech based on the concept of irrelevant variability normalization. Evaluation results on Aurora2 connected digits database demonstrate that the proposed approach achieves significant improvements in recognition accuracy compared to the traditional VTS-based HMM compensation approach.

Index Terms— Robust speech recognition, model compensation, vector Taylor series, hidden Markov model.

1. Introduction

Most of current automatic speech recognition (ASR) systems use MFCCs (Mel-Frequency Cepstral Coefficients) and their derivatives (e.g., delta, Δ, and delta-delta, ΔΔ) as speech features, and a set of Gaussian mixture continuous density HMMs (CDHMMs) for modeling basic speech units (e.g., [10]). It is well known that the performance of such an ASR system trained with clean speech will degrade significantly when the testing speech is corrupted by additive noises and convolutional distortions. To cope with the above robust ASR problem, one of the approaches proposed previously is to compensate for CDHMM parameters based on Vector Taylor Series (VTS) approximation of an explicit model of environmental distortions whose parameters can be estimated from the speech utterance to be recognized (e.g., [8, 9, 6, 1]).

In the traditional VTS-based HMM compensation approach, the set of generic HMMs are typically trained from “clean” speech only. In this paper, we present a maximum likelihood approach to training generic HMMs from both “clean” and “corrupted” speech based on the concept of irrelevant variability normalization (IVN) (e.g., [4, 5]). An independent work and "corrupted" speech based on the concept of irrelevant variability normalization. Evaluation results on Aurora2 connected digits database demonstrate that the proposed approach achieves significant improvements in recognition accuracy compared to the traditional VTS-based HMM compensation approach.

2. Overview of Traditional VTS-based HMM Compensation Approach

2.1. Notations, Distortion Function, VTS Approximation

Let’s assume that in our speech recognizer, the “clean” speech $X$ of each basic speech unit is modeled by a generic CDHMM, whose model parameters are denoted as $\lambda = \{\pi, a_{sr}, c_{sm}, \mu_{s,m}, \Sigma_{s,m}, s, s' = 1, \ldots, N_s; m = 1, \ldots, N_m\}$, where $N_s$ is the number of states, $N_m$ is the number of Gaussian components for each state, $\{\pi_s\}$ is the initial state distribution, $a_{sr}$'s are state transition probabilities, $c_{sm}$, $\mu_{s,m}$ and $\Sigma_{s,m}$ denote the mixture weights, mean vector and diagonal covariance matrix of $m$-th Gaussian component in the $s$-th state, respectively. Consequently, we use $\Lambda = \{\lambda\}$ to denote the set of generic CDHMM parameters.

Let’s also assume that in the time domain, the “corrupted” speech $y[t]$ is subject to the following distortion model

$$y[t] = h[t] * x[t] + n[t]$$

where three independent signals $x[t], h[t]$ and $n[t]$ represent the $t$-th sample of clean speech, the convolutional (e.g., channel) distortion and the additive noise, respectively. In the MFCC domain, the distortion model can be expressed approximately as

$$Y = C \log[\exp(C^{-1}(X + H)) + \exp(C^{-1}N)]$$

where $C$ and $C^{-1}$ denote the discrete cosine transform (DCT) matrix and its inverse, the log and exp functions operate at an element level on the corresponding vectors, $N$ has a Gaussian probability density function (PDF) with mean vector $\mu_n$ and covariance matrix $\Sigma_n$, $H$ has a PDF of the Kronecker delta function $\delta(H - h)$. However, in most of current ASR systems, the dimension, $D_1$, of MFCC feature vector is different from the number, $D_2$, of Mel-frequency filter banks, therefore the following distortion model has to be used in practice:

$$Y = f(X, N, H) = C_1 \log[\exp(C_1^{+1}(X + H)) + \exp(C_1^{+}N)]$$

where $C_1$ is a truncated version of the DCT matrix $C$ as

$$C_1 = [I_{D_1} \times D_1 \ 0_{D_1} \times (D_2 - D_1)] C_1$$

This work was done while Y. Hu worked at The University of Hong Kong, and was supported by grants from the RGC of the Hong Kong SAR and Anhui USTC iFLYTEK Co. Ltd., Hefei, China. Mr. Jun Du's contribution in conducting several experiments reported in this paper is also gratefully acknowledged.

In Section 2, we give an overview of the formulation of the traditional VTS-based HMM compensation approach and introduce the necessary notations. In Section 3, we present the formulation of our proposed IVN-based HMM training approach for VTS-based HMM compensation. In Section 4, we report the experimental results, and finally we conclude the paper in Section 5.
and $C^+_1$ is the Moore-Penrose inverse of $C_1$ as
\[
C^+_1 = C^{-1} \begin{bmatrix} I_{D_1 \times D_1} & 0 \end{bmatrix}_{(D_2-D_1) \times D_1}.
\]
The above nonlinear distortion function can be approximated by a linear function based on the first-order-approximation of the VTS of the original function expanded at a fixed point $(x_0, n_0, h_0)$.

Let's further assume that the abovementioned distortions will not alter the alignment between a speech frame and the corresponding Gaussian component of CDHMMs. Consequently, only the mean vector and covariance matrix, $\mu_{y, sm}$ and $\Sigma_{y, sm}$, for each Gaussian component of the "corrupted" speech CDHMM will be affected, while other CDHMM parameters remain unchanged.

Given an estimation of the distortion model parameters from the speech utterance to be recognized, $\mu_{y, sm}$ and $\Sigma_{y, sm}$ for static MFCCs can be compensated as follows:
\[
\begin{align*}
\mu_{y, sm} &= f(x_0, n_0, h_0) + A_{sm}(\mu_{x, sm} - x_0) \\
&\quad + A_{sm}(h - h_0) + (I - A_{sm})(\mu_0 - n_0) \\
\Sigma_{y, sm} &= A_{sm}\Sigma_{x, sm}A_{sm}^T + (I - A_{sm})\Sigma_\mu(I - A_{sm})^T
\end{align*}
\] (4)

where $A_{sm} = C_1B_{sm}C_1^+$ and $B_{sm}$ is a diagonal matrix whose diagonal elements are given by $\eta_{sm}[j]$ with
\[
\eta_{sm}[j] = \frac{\exp(C^+_{1}(x_0[j] + C_{1}^n h_0[j]))}{\exp(C^+_{1}(n_0[j])}.
\]

In the above equations, $C_{1}^+ x_0[j], C_{1}^n h_0[j]$ and $C_{1}^+ n_0[j]$ represent the $j$-th element of the vectors $C_{1}^+ x_0, C_{1}^n h_0$ and $C_{1}^+ n_0$ respectively; $(\cdot)^T$ denotes the transpose of a vector or a matrix.

In the traditional VTS-based HMM compensation approach, the set of generic CDHMMs are trained from "clean" speech only. In this case, we adopt a heuristic approach described in [1] to compensate for CDHMM mean parameters corresponding to dynamic features.

2.2. Robust ASR Using VTS-based HMM Compensation

Given the set of pre-trained generic CDHMMs, an unknown utterance $Y$ can be recognized by using the following procedure:

**Step 1:** Derive an ML estimation of $\mu_h, \Sigma_h$ from the first several (10 in our experiments) frames of $Y$, and set $h$ as a zero vector. Set the VTS expansion point as $x_0 = \mu_{x, sm}, n_0 = \mu_0, h_0 = h$.

**Step 2:** Given the fixed VTS expansion point and the updated estimation of the distortion model parameters, compensate CDHMMs as described above and recognize $Y$ to obtain the transcription $W_Y$.

**Step 3:** Update the distortion model parameters, $\mu_h, \Sigma_h$ and $h$ by running one EM iteration to increase the likelihood function defined based on the given fixed VTS expansion point, $Y$, and $W_Y$ [6].

**Step 4:** Update the VTS expansion point as $x_0 = \mu_{x, sm}, n_0 = \mu_0, h_0 = h$.

**Step 5:** Repeat Step 3 and Step 4 $N_{VTS}^{(i)}$ times.

**Step 6:** Repeat Step 2 to obtain the final recognition result $W_Y$.

3. IVN-based HMM Training For VTS-based HMM Compensation

Suppose we are given a set of multi-condition training data $\mathcal{Y} = \{Y_i\}_{i=1}^N$, where $Y_i$ is a sequence of $T_i, feature vectors, \mathcal{Y}$ may consist of both "clean" and "corrupted" speech. Let’s assume that for each sentence $Y_i$, there is a unique distortion model, whose parameters are denoted as $\Phi(i) = \{\mu(i), \Sigma(i), h(i)\}$, and the distortion models for different training sentences are independent. Let’s use $\Phi = \{\Phi(i)\}_{i=1}^N$ to denote the set of all distortion model parameters. Our IVN-based ML training approach is to maximize, by adjusting the generic CDHMM parameters $\Lambda$ and the distortion model parameters $\Phi$, the following likelihood function
\[
L(\Lambda, \Phi) = \prod_{i=1}^N p(Y_i|\Lambda, \Phi).
\]

In the following subsection, we describe in detail an approximate ML training procedure to solve the above problem.

3.1. ML Training Procedure

Our ML training procedure is as follows:

**Step 1:** Initialization

First, a set of CDHMMs $\Lambda$ are trained from multi-condition training data $\mathcal{Y}$. Each CDHMM state has $N_\mu$ Gaussian components for $\Lambda$. For each training sentence $Y_i$ in $\mathcal{Y}$, derive an ML estimation of $\mu(i), \Sigma(i)$ from the first several (10 in our experiments) frames of $Y_i$, and set $h(i)$ as a zero vector.

**Step 2:** Repeating Step 3 and Step 4 $N_{IVN} \times N_{Y}$ times.

**Step 3:** Estimating distortion model parameters $\Phi$

Given $\Lambda, \Phi$ can be updated to increase the likelihood function $L(\Lambda, \Phi)$ as follows:

**Step 3-1:** For each training sentence $Y_i$, update the VTS expansion point as $x^{(i)}_0 = \mu_{x, sm}, n^{(i)}_0 = \mu_0, h^{(i)}_0 = h^{(i)}$. Then, evaluate the likelihood function $L(\Lambda, \Phi)$.

**Step 3-2:** Update the distortion model parameters, $\Phi$, by running one EM iteration to increase the likelihood function defined based on the given fixed VTS expansion points and the training data $\mathcal{Y}$ as follows [6]:

**Step 4-1:**

\[
\tilde{\mu}^{(i)}_h = \frac{\sum_{t, s, m} c^{(i)}_{sm}(t) \mu^{(i)}_{x, sm}(t)}{\sum_{t, s, m} c^{(i)}_{sm}(t)}
\] (7)

**Step 4-2:**

\[
\tilde{\Sigma}^{(i)}_h = \sum_{t, s, m} c^{(i)}_{sm}(t)(\Sigma^{(i)}_{x, sm} + \mu^{(i)}_{x, sm}(t)\mu^{(i)}_{x, sm}(t))^T
\] (8)

\[
\tilde{h}^{(i)} = \left[ \sum_{t, s, m} c^{(i)}_{sm}(t)(A^{(i)}_{sm})^T(\Sigma^{(i)}_{y, sm, sm})^{-1}A^{(i)}_{sm}\right]^{-1}
\]

\[
\sum_{t, s, m} c^{(i)}_{sm}(t)(A^{(i)}_{sm})^T(\Sigma^{(i)}_{y, sm, sm})^{-1}(\mu^{(i)}_{y, sm, sm}(t) - \mu^{(i)}_{y, sm, sm}(t))
\] (9)
where

\[ \gamma_i^{(t)}(s, m) = P(s_t = s, m_t = m | Y_i, \Lambda, \Phi) \]  

(10)

\[ \eta_i^{(t)} = \mu_i^{(t)} + \Sigma_i^{(t)}(\Sigma_i^{(t)})^{-1}(y_i^{(t)} - \mu_i^{(t)}) \]  

(11)

\[ \xi_i^{(t)} = \sum_i - \eta_i^{(t)} - \psi_i^{(t)}(\Sigma_i^{(t)})^{-1}(1 - \xi_i^{(t)}) \]  

(12)

\[ \nu_i^{(t)} = \gamma_i^{(t)} - \gamma_i^{(t)} \]  

(13)

\[ \tau_i^{(t)} = (\psi_i^{(t)}(\Sigma_i^{(t)}))^{-1} = (I + A_i^{(t)})^{-1} \]  

(14)

In the above equations, we use \( \sum_{s, m = 1} \) to denote \( \sum_{s, m = 1} \sum_{m = 1} \) for notational simplicity.

**Step 3-3:** For each training sentence \( Y_i \), update the VTS expansion point as \( \bar{x}_i^{(t)} = \mu_{k_i^{(t)}} \), \( n_i^{(t)} = \bar{\mu}_i^{(t)} \), \( h_i^{(t)} = \bar{\Sigma}_i^{(t)} \). Then, evaluate the likelihood function \( \mathcal{L}(\Lambda, \Phi) \). If \( \mathcal{L}(\Lambda, \Phi) \leq \mathcal{L}(\Lambda, \Phi) \), goto **Step 4**; Otherwise, goto **Step 3-4**.

**Step 3-4:** Repeat **Step 3-2** and **Step 3-3** \( N_{VT}^{(2)} \) times.

**Step 4:** Estimating generic CDHMM parameters \( \Lambda \)

Given \( \Phi \), \( \Lambda \) can be updated to increase the likelihood function \( \mathcal{L}(\Lambda, \Phi) \) as follows:

**Step 4-1:** For each training sentence \( Y_i \), update the VTS expansion point as \( \bar{x}_i^{(t)} = \mu_{k_i^{(t)}} \), \( n_i^{(t)} = \bar{\mu}_i^{(t)} \), \( h_i^{(t)} = \bar{\Sigma}_i^{(t)} \). Then, evaluate the likelihood function \( \mathcal{L}(\Lambda, \Phi) \).

**Step 4-2:** Update the generic CDHMM parameters, \( \Lambda \), by running one EM iteration to increase the likelihood function defined based on the fixed VTS expansion points and the training data \( \mathcal{D} \) as follows:

\[ \mu_{x_i} = \sum_{i=1}^{I} \gamma_{i}^{(t)}(1)/I \]  

(15)

\[ \bar{\alpha}_{x, s} = \sum_{i=1}^{I} \sum_{t=1}^{T} \xi_{i}^{(t)}(s, m) \]  

(16)

\[ \bar{\zeta}_{x, s} = \sum_{i=1}^{I} \sum_{t=1}^{T} \xi_{i}^{(t)} \]  

(17)

\[ \bar{\nu}_{x, s} = \sum_{i=1}^{I} \sum_{t=1}^{T} \gamma_{i}^{(t)} \]  

(18)

\[ \bar{\Sigma}_{x, s} = \sum_{i=1}^{I} \sum_{t=1}^{T} \zeta_{i}^{(t)} \]  

(19)

where

\[ \gamma_{i}^{(t)} = P(s_{t} = s | Y_t, \Phi, \Lambda) \]  

(20)

\[ \xi_{i}^{(t)} = P(s_{t-1} = s, s_{t} = s | Y_t, \Phi, \Lambda) \]  

(21)

\[ \xi_{i}^{(t)} = P(s_{t} = s, Y_t) - \mu_{i}^{(t)} \]  

(22)

\[ \Sigma_{x, s}^{(t)} = \sum_{i=1}^{I} \sum_{t=1}^{T} \xi_{i}^{(t)} \]  

(23)

\[ \Sigma_{x, s}^{(t)} = \sum_{i=1}^{I} \sum_{t=1}^{T} \xi_{i}^{(t)} \]  

(24)

\[ \tau_{i}^{(t)} = (\psi_{i}^{(t)}(\Sigma_{i}^{(t)}))^{-1} = (I + A_{i}^{(t)})^{-1} \]  

(25)

**Step 4-3:** For each training sentence \( Y_i \), update the VTS expansion point as \( \bar{x}_i^{(t)} = \mu_{k_i^{(t)}} \), \( n_i^{(t)} = \bar{\mu}_i^{(t)} \), \( h_i^{(t)} = \bar{\Sigma}_i^{(t)} \). Then, evaluate the likelihood function \( \mathcal{L}(\Lambda, \Phi) \). If \( \mathcal{L}(\Lambda, \Phi) \leq \mathcal{L}(\Lambda, \Phi) \), goto **Step 2**; Otherwise, goto **Step 4-4**.

**Step 4-4:** Repeat **Step 4-2** and **Step 4-3** \( N_{VT}^{(3)} \) times.

### 3.2. Discussions

In the above IVN-based training procedure, the VTS-based CDHMM compensation only applies to the means and covariance matrices corresponding to the static MFCC features. The CDHMM means and covariance matrices corresponding to the dynamic features are the ones that are trained from the IVN-based training procedure. This is also true for the robust ASR procedure described in Section 2.2 when the IVN-trained models are used as generic CDHMMs, which is different from traditional VTS-based CDHMM compensation approach as described in Section 2.1.

It is also noted that the updating formulas in Eq. (18) and Eq. (19) are for static MFCC features only. The updating formulas for dynamic features are similar to the traditional ones for ML training of CDHMM with a new \( \zeta_{i}^{(t)}(t) \) as defined in Eq. (10).

### 4. Experiments and Results

#### 4.1. Experimental Setup

In order to verify the effectiveness of the proposed IVN-based CDHMM training method and compare it to the traditional VTS-based CDHMM compensation approaches [8, 9, 6, 1], a series of experiments are performed for the task of speaker independent recognition of connected digit strings on Aurora2 database. A full description of the Aurora2 database and a test framework is given in [2].

In our ASR systems, the feature vector we used consists of 13 (i.e., \( D_1 = 13 \)) MFCCs (including \( C_0 \)) plus their first and second order derivatives. The speech data is processed in a time window of 25ms, shifted every 10ms. A pre-emphasis with a coefficient of 0.97 is performed. The number of Mel-frequency filter banks is 23 (i.e., \( D_2 = 23 \)). The cepstra are computed based on the power spectral density. The delta and delta-delta features are extracted using linear regression method as detailed in [10] with a setting of relevant parameters in HTK notations as \( deltawindow = 3 \) and \( accwindow = 2 \). Each digit is modeled by a whole word left-to-right CDHMM, which consists of 18 states, each having 20 Gaussian mixture components (i.e., \( N_m = 20 \)) with diagonal covariance matrices. Besides, two pause models, "sil" and "sp", are created to model the silence before/after the digit string and the short pause between any two digits.

Two baseline systems are trained first. In the first system (referred to as Baseline-CT hereinafter), CDHMMs are trained from 8440 clean speech sentences (referred to as CT condition in [2]). In the second system (referred to as Baseline-MT hereinafter), CDHMMs are trained from 8440 sentences that come from 20 subsets representing 4 different noise scenarios (i.e., \( D_1 = 13 \)) with a setting of relevant parameters in HTK notations as \( deltawindow = 3 \) and \( accwindow = 2 \). Each digit is modeled by a whole word left-to-right CDHMM, which consists of 18 states, each having 20 Gaussian mixture components (i.e., \( N_m = 20 \)) with diagonal covariance matrices. Besides, two pause models, "sil" and "sp", are created to model the silence before/after the digit string and the short pause between any two digits.
Table 1: Performance (word accuracy in %) comparison of two baseline systems and three robust ASR systems using VTS-based CDHMM compensation. The performance is averaged over SNRs between 0 and 20 dB on three different test sets of Aurora2 database.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Set A</th>
<th>Set B</th>
<th>Set C</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline-CT</td>
<td>62.88</td>
<td>56.78</td>
<td>74.15</td>
<td>62.09</td>
</tr>
<tr>
<td>Baseline-MT</td>
<td>91.50</td>
<td>89.54</td>
<td>89.19</td>
<td>90.25</td>
</tr>
<tr>
<td>Clean-VTS</td>
<td>87.78</td>
<td>87.63</td>
<td>87.89</td>
<td>87.74</td>
</tr>
<tr>
<td>MT-VTS</td>
<td>92.73</td>
<td>91.24</td>
<td>93.05</td>
<td>92.20</td>
</tr>
<tr>
<td>IVN-VTS</td>
<td>93.36</td>
<td>92.86</td>
<td>93.05</td>
<td>93.10</td>
</tr>
</tbody>
</table>

Table 2: Performance (word accuracy in %) comparison of several methods averaged over three test sets of Aurora2 database at each SNR.

<table>
<thead>
<tr>
<th>Methods</th>
<th>0dB</th>
<th>5dB</th>
<th>10dB</th>
<th>15dB</th>
<th>20dB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline-MT</td>
<td>68.76</td>
<td>89.87</td>
<td>93.91</td>
<td>97.93</td>
<td>98.79</td>
</tr>
<tr>
<td>Clean-VTS</td>
<td>63.57</td>
<td>84.65</td>
<td>93.95</td>
<td>97.68</td>
<td>98.87</td>
</tr>
<tr>
<td>MT-VTS</td>
<td>76.12</td>
<td>91.71</td>
<td>96.27</td>
<td>98.06</td>
<td>98.81</td>
</tr>
<tr>
<td>IVN-VTS</td>
<td>77.50</td>
<td>92.73</td>
<td>97.23</td>
<td>98.82</td>
<td>99.21</td>
</tr>
</tbody>
</table>

other 4 types of noises (i.e., restaurant, street, airport and train station) are added to its subset with also 7 SNRs. For the Test Set C, suburban train and street noises are used as the additive noise sources but the speech and noise are filtered with a MIRS characteristic while the G.712 characteristic is used in multi-condition training set as well as the first two test sets. Both the training of the above baseline systems and the recognition experiments were performed by using the HTK [10] and the standard scripts provided by ETSI [2].

In IVN-based CDHMM training, the set of initial CDHMMs described in Step 1 are trained from the same set of multi-condition switch data for training the Baseline-MT system. The relevant control parameters are set as $N_{VTS}^{(1)} = 5$, $N_{VTS}^{(2)} = 4$, $N_{IVN} = 1$.

4.2. Experimental Results

Table 1 summarizes a performance (word accuracy in %) comparison of two baseline systems and three robust ASR systems using VTS-based CDHMM compensation with three different sets of generic CDHMMs, namely CDHMMs from “Baseline-CT” system (labeled as “Clean-VTS”), CDHMMs from “Baseline-MT” system (labeled as “MT-VTS”), and CDHMMs trained from multi-condition training data using IVN-based training proposed in this paper (labeled as “IVN-VTS”), respectively. The robust ASR procedure described in Section 2.2 is used for VTS-liked experiments with the control parameter set as $N_{IVN}^{(1)} = 5$. The performance is averaged over SNRs between 0 and 20 dB on test Set A, Set B and Set C respectively. It is observed that “Clean-VTS” performs better than “Baseline-CT” and “MT-VTS” performs better than “Baseline-MT”. This confirms the effectiveness of the VTS-based HMM compensation approaches. However, “Clean-VTS” performs worse than “Baseline-MT” and “MT-VTS”, which implies that it is important to have a set of appropriate generic CDHMMs for taking full advantage of the potential offered by VTS-based HMM compensation approaches. This is confirmed by the observation that our proposed IVN-VTS method achieves much better performance in all the test sets than that of both the traditional VTS-based HMM compensation approaches (“Clean-VTS” and “MT-VTS”) and two baseline systems (“Baseline-CT” and “Baseline-MT”). Table 2 gives a performance (word accuracy in %) comparison of several methods averaged over three test sets of Aurora2 database at each SNR (in dB). Similar observations are made under different SNRs. Overall, the proposed IVN-VTS approach achieves the best performance in all testing conditions.

5. Conclusion and Discussions

In this paper, we have presented a maximum likelihood approach to training generic CDHMMs from multi-condition speech data based on the concept of irrelevant variability normalization and the VTS approximation of an explicit model of environmental distortions. Its effectiveness has been confirmed in an experimental study on both Aurora2 and Aurora3 databases, but only experimental results on Aurora2 are reported in this paper due to the limit of paper length. Ongoing and future works include 1) to explore the similar idea for another type of HMM compensation approach using Unscented Transformation (UT) [3], and 2) to compare the above approaches with other types of robust ASR approaches (e.g., [5, 4]). We will report those results elsewhere when they become available.

6. References