On Noise Tracking for Noise Floor Estimation*

Mahdi Triki
Philips Research, Eindhoven, The Netherlands
mahdi.triki@philips.com

Abstract

Various speech enhancement techniques (e.g. noise suppression, dereverberation) rely on the knowledge of the statistics of the clean signal and the noise process. In practice, however, these statistics are not explicitly available, and the overall enhancement accuracy critically depends on the estimation quality of the unknown statistics. With this respect, subspace based approaches have shown to allow for reduced estimation delay and perform a good tracking vs. final misadjustment tradeoff [3, 4]. For an accurate noise non-stationarity tracking, these schemes have the challenge to estimate the correlation matrix of the observed signal from a limited number of samples. In this paper, we investigate the effect of the covariance estimation artifacts on the noise PSD tracking. We show that the estimation downsides could be alleviated using an appropriate selection scheme.

Index Terms: speech enhancement, noise floor, non-stationary noise, median filter, subspace methods

1. Introduction

Speech enhancement aims at improving the performance of audio communication in a noisy environment. Several practical methods have already been proposed. Among them, the class of frequency domain methods has been relatively successful due to their implementation simplicity and their capability of handling noise non-stationarity to some extent. These schemes recover the clean signal by applying an appropriate gain filter. The design of these filters relies on the knowledge of the clean and noise signal statistics. In practice however, these statistics are not explicitly available and should be estimated. The accuracy of the overall enhancement approach critically depends on the estimation quality of the unknown statistics.

Joint clean speech and noise Power Spectral Density (PSD) estimation is an underdetermined problem. In fact, using a unique observation, we aim tracking both clean speech and noise statistics. A classic trick to overcome the underdeterminacy problem is to exploit speech pauses. The basic observation is that the speech signal is not always present. Then, the noise PSD can be estimated and updated during the speech pauses. The noise is typically updated either using a Voice Activity Detector (to identify speech pause periods) [1] or by tracking the minimum statistics (MS) of the noisy signal [2]. Recently, a subspace decomposition based scheme was proposed for noise floor estimation [3]. The subspace considered herein characterizes the time evolution of the noisy Discrete Fourier Transform (DFT) coefficients. The key observation is that in such a domain, the speech signal can be described with a low rank model[4], while the noise is rather full rank[6]. Therefore, a noise subspace can be identified, and the noise PSD is still updated even if speech is constantly present [3, 4]. Due to space limitation, no detailed overview on noise subspace estimation and tracking will be presented in this paper. The interested reader is referred to [4, 5, 6] for an exhaustive description and better coverage.

Subspace based techniques track noise subspace in DFT domain. For every frequency bin f, the DFT coefficients $y(n,f)$ of the received signal are organized in a $K \times N$ Hankel matrix $\mathbf{Y}(n,f)$:

$$\{\mathbf{Y}(n,f)\}_{i,j} = y(n-i-j+N_c, f) \quad \begin{array}{c} i = 0:(K-1) \\ j = 0:(N-1) \end{array}$$

where K denotes the covariance matrix size, N is the sample support size, and N_c characterizes (an eventual) non-causal estimation. The sample covariance matrix (at the time frame n and the frequency bin f) is computed as

$$\widehat{\mathbf{R}}_y(n,f) = \frac{1}{N}\,\mathbf{Y}(n,f)\mathbf{Y}(n,f)^H \tag{1}$$

where $(.)^H$ represents the complex-conjugate (Hermitian) transpose operator.

Subspace based techniques hinges on the observation that speech DFT coefficients live often in a low-dimension subspace. Thus, the smallest eigenvalue of the *DFT covariance matrix* $\lambda_{y,min}(n,f)$ provides *often* a consistent information on the noise PSD. The Minimum Subspace Noise Tracking (MSNT) scheme exploits these facts and operates in two steps [4]:

1. *Estimate* the best representative of the noise subspace (i.e., $\lambda_{y,min}(n,f)$), at each time-frequency bin (n,f).
2. *Select* the best candidate to update the noise PSD using minimum search (in analogy with the MS tracking), i.e.,

$$\widehat{\sigma}_v^2(n,f) = \frac{1}{B}\min_{i=0:M-1}\{\lambda_{y,min}(n-i,f)\} \tag{2}$$

where B denotes a bias compensation factor, and M characterizes the search memory.

In practice, the scheme has the extra challenge to estimate the DFT covariance matrix from a limited number of samples, to allow fast tracking of noise non-stationarity. To assess the quality of noise floor estimation, a stationary white noise is synthetically added to a speech signal (8 kHz sampling frequency). Next, $\hat{\lambda}_{y,min}$ were estimated over consecutive M frame, sorted (in ascendant order), and assessed using the Flatness Measure (defined and discussed in [4]). Averaged over frequencies, the FM measures the non-stationarity of the *sorted* smallest eigenvalues: the closer to one, the more stationary is the quantity, and the better the noise estimated (the lower the (bursty) speech leakage). Figure 1. plots the flatness measure (y-axis) of $\left\langle \hat{\lambda}_{y,min}(n) \right\rangle_f = \sum_f \hat{\lambda}_{y,min}(n,f)$, sorted over the last $M = 20$ frames (x-axis). The quantities are estimated using $N = K = 7$.

*The herein disclosed information is secret until august 28^{th} 2011; eyes only for the appointed reviewers of this conference.

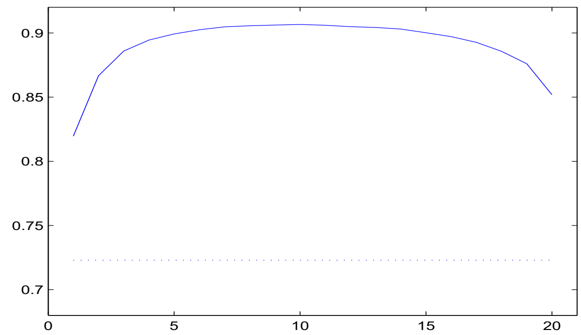

Figure 1: Flatness of $\left\langle \hat{\lambda}_{y,min} \right\rangle_f$, estimated using $N = K = 7$, and sorted over the last $M = 20$ frames.

Curves show a drop in the FM both in the large and low magnitude (very right and left) regions. These distortions were associated with the speech leakage and the sample covariance estimation, respectively. Curves also reveal that (for $N \approx K$), using a minimum search approach (as in (2)) is not the best option. In [7], we have proposed an alternative median-search approach, and demonstrated an enhancement in noise PSD estimation and auditive results. However, we have noticed that the median-search approach is less robust to speakers variabilities.

In the present paper, we investigate and justify (using random matrix theory) the behavior observed in Figure 1. We will also optimize the selection scheme to allow better accuracy/robustness tradeoff.

Notations: Upper- and lower-case boldface letters denote matrices and vectors, respectively. Upper- and lower-case normal letters represent scalar constants and processes, respectively. Either as a subscript or as an argument n and f refer respectively to the time-frame, and frequency-bin indices.

2. Random Matrix Theory for Noise Floor Estimation

In this section, we review some results from random matrix theory, and their implications on subspace-based noise tracking. The random matrix theory has seen exciting recent development, and provided useful analysis tools that have shown to be insightful in several signal processing applications (e.g. wireless communication, financial data analysis) [8]. Some of these tools were applied for covariance matrix analysis, and gave useful information even for small dimensions. Many of these analysis and results were derived under the assumption that the columns of $\mathbf{Y}$ are independent Gaussian (with a covariance matrix $\mathbf{\Sigma}$)[1] . In such a case the sample covariance matrix $\widehat{\mathbf{R}}_y = \frac{1}{N}\mathbf{Y}\mathbf{Y}^H$ is said to have a Wishart distribution $\widehat{\mathbf{R}}_y \sim \mathcal{W}_K(N, \mathbf{\Sigma})$ (with a dimension K and N degree of freedom). The special case $\mathbf{\Sigma} = \mathbf{I}$ (K-dimensional identity matrix) is of a particular interest, as noise covariance matrix is assumed spherical.

The usual 'large sample' framework assumes that N/K, the number of observations per variable, is large. Such an assumption will alter the tracking capability of the noise nonstationarity: we are rather interested in the case $N/K \to 1^+$. Under such an assumption, a basic phenomenon is that the sample eigenvalues are more spread out than theory. Consider one random observation of a (real value) Wishart matrix $\sim \mathcal{W}_7(7, \mathbf{I})$. Its ordered sample eigenvalue were

$$\hat{\lambda}_i = 0.005,\ 0.077,\ 0.16,\ 0.67,\ 1.09,\ 1.82,\ 2.76$$

In this case, the ratio of largest to smallest is about 500! Investigating the eigenvalue statistical properties, two general areas are distinguished:

- the bulk: refers to the properties of the full eigenvalues set $\hat{\lambda}_1 \leq \hat{\lambda}_2 \leq \ \leq \hat{\lambda}_K$.
- the extremes: addresses the largest and smallest eigenvalues.

2.1. Bulk spectrum

In [9], Marchenko and Pastur investigated the empirical distribution of the eigenvalues of a Wishart matrix $\sim \mathcal{W}_K(N, \mathbf{I})$[2]. They showed that if both N and K tend to ∞, in some given ratio $\frac{N}{K} \to \gamma$,

$$\frac{1}{K}\#\left\{\hat{\lambda}_i : \hat{\lambda}_i < x\right\} \to F(x) \tag{3}$$

where:
- $F^{'}(x) = \dfrac{1}{2\pi x\gamma}\ \sqrt{(b-x)(x-a)}a < x < b$ is the limiting density (Figure 2).
- $a = \left(1 - \sqrt{\gamma}\right)^2$ and $b = \left(1 + \sqrt{\gamma}\right)^2$ characterize the (non-null) support of the limiting density.

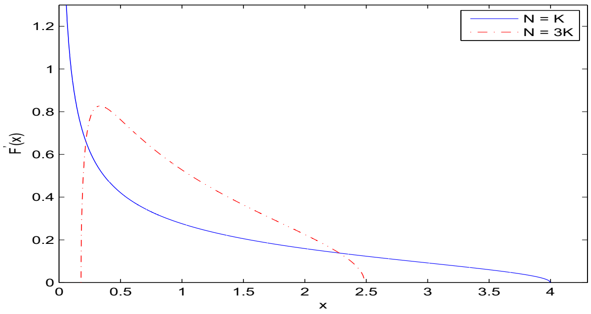

Figure 2: Limiting distribution density for $\gamma = 1$ (solid line) and $\gamma = 3$ (dotted line).

Moreover, there are no stray eigenvalues in the sense that the largest and smallest eigenvalues converges to the edges of the support of $F(x)$. Specifically, the smallest eigenvalue converge almost surely to [11]

$$\hat{\lambda}_{min} \to (1 - \sqrt{\gamma})^2 \tag{4}$$

These results justify the eigenvalue spread illustrated in the example before, and prove that such spread do not disappear even asymptotically. The spread of the sampled eigenvalue (and the quality of the eigenvalue estimation) exclusively depends on the ratio $\gamma = N/K$: the smaller γ, the more spread, and the noisier estimation. Moreover, when $\gamma \to 1$, the limit of $\hat{\lambda}_{min}$ (which correspond to the bias compensation factor B in (2)) gets dangerously close to 0.

2.2. smallest Eigenvalue

We consider now the left-hand edge, namely the smallest eigenvalue. It has been observed that tracking the smallest eigenvalue

[1] The results presented hereafter hold (with a good approximation) for a Hankel matrix $\mathbf{Y}$ with i.i.d. Gaussian entries (as assumed in our problem statement).

[2] The previous result was extended to the case where $\mathbf{\Sigma} = \mathbf{I} + \mathbf{\Sigma}_r$ and $\mathbf{\Sigma}_r$ has a fixed limited rank r [10].

of DFT covariance matrix provides a consistent information regarding the noise PSD [4]. The smallest eigenvalue plays also a central role in many other applications, for instance in determining the error rate in a MIMO transmission [12].

The distribution of the smallest eigenvalue was first derived in terms of matrix determinants, whose elements contain incomplete gamma functions and Nuttall Q-functions (e.g. [12]). These expressions were intractable and not suitable for further mathematical analysis (e.g. integration). Later, recursive solution [13] and polynomial approximation [14] were derived.

Since the distribution for the smallest eigenvalue of the general Wishart matrix is too complicated, special Wishart matrices are studied with the purpose of developing concise expressions. An interesting case (in the scope of this work) is when $N \approx K$. With this respect, Forster investigated the case N large and $N - K$ fixed [15]. Particularly, he showed that the smallest eigenvalue of a Wishart Matrix $\sim \mathcal{W}_K(K+1, \mathbf{I})$ was of exponential distribution. This results is also valid for an arbitrary K (not only asymptotically) [13], and in case the columns (or rows) of the matrix $\mathbf{Y}$ are correlated [16].

3. Noise Tracking in Sample DFT Domain

In this section, we consider noise PSD estimation by tracking the smallest eigenvalue of the sample DFT covariance matrix of the received signal (as described in the introduction section). To enable effective tracking of the noise non-stationarity, we consider the case $N \approx K$ (i.e. $N - K \ll K$). In such a case, the density law of the smallest eigenvalue is approximated exponential

$$f_{\lambda_{min}}(x) = \frac{1}{\mu} e^{-x/\mu}, \qquad x \geq 0 \tag{5}$$

If $N = K + 1$ and the columns of $\mathbf{Y}$ are i.i.d., the model is exact (with $\mu = 1$).

The smallest eigenvalues are estimated on M consecutive time frames, and sorted in ascendant order:

$$\hat{\lambda}^{(1)}_{y,min} \leq \cdots \leq \hat{\lambda}^{(k)}_{y,min} \leq \cdots \leq \hat{\lambda}^{(M)}_{y,min} \tag{6}$$

To update the noise PSD, the original MSNT performs minimum selection (as described in (2))[4]:

$$\widehat{\sigma}^2_v(n, f) = \frac{1}{B^{(1)}} \hat{\lambda}^{(1)}_{y,min}(n - i, f) \tag{7}$$

In [7], we have shown that better results could be achieved via median selection:

$$\widehat{\sigma}^2_v(n, f) = \frac{1}{B^{(\lfloor M/2 \rfloor)}} \hat{\lambda}^{(\lfloor M/2 \rfloor)}_{y,min}(n - i, f) \tag{8}$$

Hereafter, we consider the general case: the noise PSD is updated using the k^{th}-sorted candidate, i.e.,

$$\widehat{\sigma}^2_v(n, f) = \frac{1}{B^{(k)}} \hat{\lambda}^{(k)}_{y,min}(n - i, f) \tag{9}$$

where $B^{(k)}$ is the bias compensation factor corresponding to the k^{th} candidate selection. Being unbiased, the noise PSD estimator $\widehat{\sigma}^2_v$ should be designed to minimize the estimation variance (i.e., MSE).

Under i.i.d assumptions, the distribution of the k^{th}-sorted candidate is

$$\begin{aligned} f_{\lambda^{(k)}_{min}}(x) &= \frac{1}{\mu} \frac{M!}{(k-1)!(M-k)!} e^{-\frac{M-k+1}{\mu} x} \left(1 - e^{-\frac{x}{\mu}}\right)^{k-1} \\ &= \frac{1}{\mu} \frac{M!}{(k-1)!(M-k)!} \sum_{i=0}^{k-1} \binom{k-1}{i} e^{-\frac{M-k+i+1}{\mu} x} \end{aligned}$$

Given that

$$\begin{aligned} \int_0^\infty x e^{-\alpha x} &= 1/\alpha^2 \\ \int_0^\infty x^2 e^{-\alpha x} &= 1/\alpha^3 \end{aligned}$$

The bias compensation factor $B^{(k)}$ and the Mean Squares Error $\text{MSE}^{(k)}$ of the estimator defined in (9) can be expressed as:

$$B^{(k)} = \mu \frac{M!}{(k-1)!(M-k)!} \sum_{i=0}^{k-1} \binom{k-1}{i} \frac{1}{M-k+i+1}$$

$$\text{MSE}^{(k)} = \frac{M!}{(k-1)!(M-k)!} \frac{\displaystyle\sum_{i=0}^{k-1} \binom{k-1}{i} \frac{1}{(M-k+i+1)^3}}{\left(\displaystyle\sum_{i=0}^{k-1} \binom{k-1}{i} \frac{1}{(M-k+i+1)^2}\right)^2}$$

Remark that the estimation MSE depends exclusively on the selection order k and the selection memory M, and it is independent from the exponential attribute μ. Next, we plot the MSE function of the selection order (Figure 3).

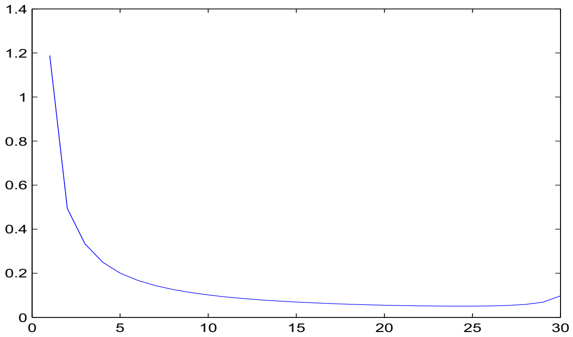

Figure 3: Mean Square Error $\text{MSE}^{(k)}$ function of the selection order k (for $M = 30$ frames).

Consistent with Figure 1, curve show that (for $N \approx K$) minimum selection (as in (2)) is not the best option; and that after a given selection order, the improvement becomes minor. The order k should be selected such to:

- achieve a good $\text{MSE}^{(k)}$: say, reaching 90% of the minimum MSE
- enhance robustness w.r.t speaker characteristics: the smallest value fulfilling the MSE requirement [7]

4. Experimental Results

Having an unbiased noise estimate is crucial for enhancement applications: an overestimation leads to over-suppression and to more signal distortion; while an underestimation leads to a high level of residual noise. The robustness of the (formerly computed) compensation factor to the speaker characteristics is then highly desired. To investigate robustness of the various MSNT varient, the bias compensation factors are learned using four different speakers (2 males and 2 females), respectively. The input SNR is 5 dB. The learning is performed as described in [4, 7]. We have plotted the bias compensation factor $B^{(k)}$function of the search memory M for:

- minimum selection [4] (Figure 4)
- median selection [7] (Figure 5)
- optimized selection, as described above (Figure 6)

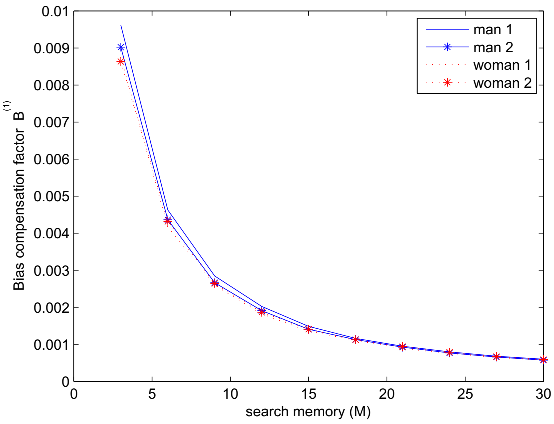

Figure 4: Robustness of the MSNT (with minimum-search tracking) bias compensation to speaker characteristics.

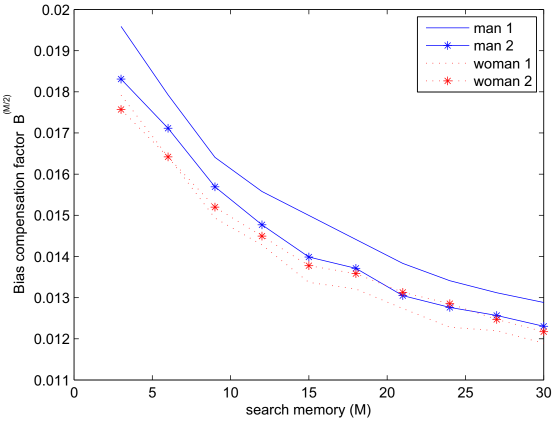

Figure 5: Robustness of the MSNT (with median-search tracking) bias compensation to speaker characteristics.

Curves show[3] that a moderate increase of the selection order (contrary to median selection) do not have considerable downsides on the robustness of the scheme w.r.t speaker characteristics, especially if the search memory M is not very small. Moreover, applying the three MSNT variants for tracking nonstationary white noise [7] shows that an appropriate selection order (as suggested before) performs comparable estimation accuracy and auditive performance with median selection approach (Table 1).

tracking method	MS	MSNT$_{min}$	MSNT$_{med}$	MSNT$_{k=6}$
MSE (in dB)	-3.82	-6.33	-6.5	-6.51

Table 1: MSE of a white non-stationary noise PSD tracking.

Remark also that (compared to median filtering approach) the optimized selection scheme has a reduced complexity, as it does not require a complete sorting of the search space.

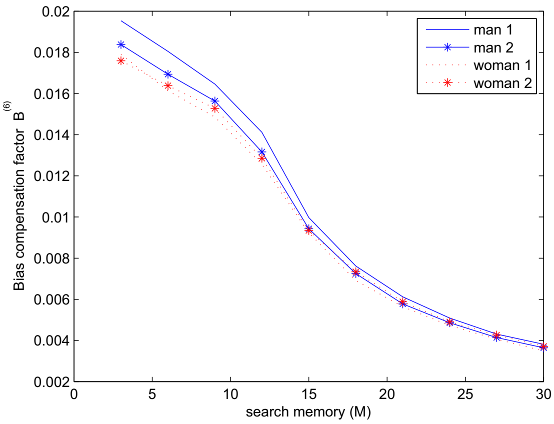

Figure 6: Robustness of the MSNT (with optimized selection order $k = 6$) bias compensation to speaker characteristics.

5. Concluding Remarks

In the present paper, we have investigated the effect of sample covariance estimation noise on the subspace-based noise tracking performance. We have demonstrated (using results derived from the random matrix theory) that if the covariance matrix and the sample support sizes are comparable, minimum tracking is not the best option. We have shown that alternative selection schemes may lead to better accuracy/robustness tradeoff.

[3] The reader should focus on the inconsistency between speakers, rather than the absolute bias curve shape (inherent to the selection scheme).

6. References

[1] S.G. Tanyer, H. Ozer, "Voice Activity Detection in Nonstationary Noise," *IEEE Trans. on Speech and Audio Processing*, Jul. 2000.

[2] R. Martin, "Noise Power Spectral Density Estimation Based on Optimal Smoothing and Minimum Statistics," *IEEE Trans. on Speech and Audio Processing*, Jul. 2001.

[3] R. C. Hendriks, J. Jensen and R. Heusdens, "Noise Tracking using DFT Domain Subspace Decompositions," *IEEE Trans. on Audio, Speech, and Language Processing*, Mar. 2008.

[4] M. Triki and K. Janse, "Minimum Subspace Noise Tracking for Noise Power Spectral Density Estimation," *In Proc of ICASSP*, Apr. 2009.

[5] M. Triki and K. Janse, "Bias Considerations for Minimum Subspace Noise Tracking," *In Proc of Interspeech*, Sep. 2010.

[6] M. Triki, "New Insights into Subspace Noise Tracking," *In Proc of Interspeech*, Sep. 2010.

[7] M. Triki,"Median Tracking in Noise Subspace for Noise Floor Estimation," *In Proc of HSCMA*, May 2011.

[8] A.M. Tulino and S. Verdu, "Random Matrix Theory and Wireless Communications," *now Publishers Inc.*, 2004.

[9] V.A. Marcenko and L.A. Pastur, "Distribution of Eigenvalues for Some Sets of Random Matrices," *Mathematics of the USSR-Sbornik*, 1967.

[10] J. Baik and J.W. Silverstein, "Eigenvalues of Large Sample Covariance Matrices of Spiked Population Models," *Journal of Multivariate Analysis archive*, Jul. 2006.

[11] J.W. Silverstein, "The Smallest Eigenvalue of a Large Dimensional Wishart Matrix," *Annals of Probability*, 1985.

[12] G. Burel, "Statistical Analysis of the Smallest Singular Value in MIMO Transmission Systems," *In Proc of ICOSSIP*, Sep. 2002.

[13] A. Edelman, "The Distribution and Moments of the Smallest Eigenvalue of a Random Matrix of Wishart Type," *Linear Algebra Appl.*, 1991.

[14] H. Zhang, F. Niu, H. Yang, X. Zhang, D. Yang, "Polynomial Expression for Distribution of the Smallest Eigenvalue of Wishart Matrices," *In Proc of VTC*, Sep. 2008.

[15] P.J. Forrester,"Painlev Transcendent Evaluation of the Scaled Distribution of the Smallest Eigenvalue in the Laguerre Orthogonal and Symplectic Ensembles," *Technical Repot*, 1991.

[16] F. Niu, H. Zhang, H. Yang, D. Yang, "Distribution of the Smallest Eigenvalue of Complex Central Semi-Correlated Wishart Matrices," *In Proc of ISIT*, Jul. 2008.