Then, such that be an truncated gaussian, we develop perceptual SURE d the (non-random) parameter that minimizes Observation model For input SNR greater than 5 dB, the proposed denoising functions.

Suppose also that Truncating the series beyond Shrinkage estimator: where Let and let f : R → R be an n-fold indefinite integral of the Lebesgue measurable function f, which is the n-th derivative of f. Suppose also that where \( f(\cdot) \). Then, \( E(Wf (W)) = \sigma^2 (f(\cdot)W\sigma^{-2}) + \sigma^2 (n-1) \langle W \rangle W^{-2}. \)

1. Overview
- We address the problem of suppressing noise from noisy speech within a risk minimization framework.
- The clean signal is estimated by minimizing an unbiased estimate of the risk function.
- We develop unbiased estimates of perceptual distortion functions.
- Minimize risk estimates to obtain the optimal denoising functions.
- For input SNR greater than 5 dB, the proposed algorithms outperform three benchmarking algorithms in terms of PESQ and SSNR scores.

2. Risk estimation principle
- Observation model: Let \( x_n = s_n + w_n \) \( n = 1, 2, \ldots, N. \)
- Parameter estimation: Obtain an estimate \( \hat{s}_n \) of the (non-random) parameter that minimizes the risk: where d measures the closeness between s and \( \hat{s}. \)
- Risk estimation approach: Since \( \mathbb{R} \) depends on s, we estimate \( \mathbb{R} \) and minimize it.
- SURE: An unbiased estimate of the MSE under i.i.d. gaussian assumption [1].
- Our contribution: Under the assumption a priori SNR is high and additive noise is a truncated gaussian, we develop perceptual risk estimates.
- Perceptual risk estimate is minimized to obtain the optimum shrinkage estimator.

3. Perceptual risk estimation
- Inakura-Saito distortion: where
- Shrinkage estimator: \( \hat{s}_n = \frac{1}{\hat{\sigma}_n} \hat{x}_n \)
- Truncating the series beyond \( n = 4 \) yields
- Generalized Stein’s Lemma: Let W be a real random variable with p.d.f

\[
p(x; c_1, c_2, \sigma) = \frac{1}{2\pi \sigma^2} \exp \left( -\frac{x^2}{2\sigma^2} \right) I_{-c_1x < c_2} \]

where \( K = \frac{1}{\sqrt{2\pi \sigma^2}} \exp \left( -\frac{u^2}{2\sigma^2} \right) du \) and let \( f : R \rightarrow R \) be an n-fold indefinite integral of the Lebesgue measurable function \( f \), which is the n-th derivative of f. Suppose also that where \( f(\cdot) \). Then, \( E(Wf (W)) = \sigma^2 (f(\cdot)W\sigma^{-2}) + \sigma^2 (n-1) \langle W \rangle W^{-2}. \)

4. Performance Comparison
Results averaged over 10 different speech files and 50 different noise realizations (NOIZEUS database)

<table>
<thead>
<tr>
<th>Risk</th>
<th>MSE</th>
<th>( \hat{\sigma}_n^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td>( \hat{b}_n^2 )</td>
<td>1 - \frac{1}{\hat{\sigma}^4}</td>
</tr>
<tr>
<td>log MSE</td>
<td>( \log \hat{b}_n^2 )</td>
<td>exp \left[ \frac{1}{\hat{\sigma}^4} \right]</td>
</tr>
<tr>
<td>WE</td>
<td>( \hat{b}_n^2 )</td>
<td>\frac{1}{\hat{\sigma}^2}</td>
</tr>
<tr>
<td>IS-II</td>
<td>( \hat{b}_n^2 )</td>
<td>\frac{1}{\hat{\sigma}^4}</td>
</tr>
<tr>
<td>COSH</td>
<td>( \hat{b}_n^2 )</td>
<td>\frac{1}{\hat{\sigma}^4}</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Risk</th>
<th>MSE</th>
<th>( \hat{\sigma}_n^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td>( \hat{b}_n^2 )</td>
<td>1 - \frac{1}{\hat{\sigma}^4}</td>
</tr>
<tr>
<td>log MSE</td>
<td>( \log \hat{b}_n^2 )</td>
<td>exp \left[ \frac{1}{\hat{\sigma}^4} \right]</td>
</tr>
<tr>
<td>WE</td>
<td>( \hat{b}_n^2 )</td>
<td>\frac{1}{\hat{\sigma}^2}</td>
</tr>
<tr>
<td>IS-II</td>
<td>( \hat{b}_n^2 )</td>
<td>\frac{1}{\hat{\sigma}^4}</td>
</tr>
<tr>
<td>COSH</td>
<td>( \hat{b}_n^2 )</td>
<td>\frac{1}{\hat{\sigma}^4}</td>
</tr>
</tbody>
</table>

Implementation details:
- We apply shrinkage estimator in DCT domain.
- Framework processing: Frame length = 40 ms, 75% Overlap, \( Fs = 8 \) kHz.
- Benchmarking denoising algorithms: WFIL [3], LMSE [4], and BNMF [5].

5. Conclusion
- Introduced the notion of risk estimation for single-channel speech enhancement.
- Proposed unbiased estimates for perceptual distortion functions.
- Minimize risk estimates to obtain the optimum denoising functions.
- For SNR greater than 5 dB, the proposed approach resulted in better denoising performance than the benchmarking techniques.

6. References
PROSE: Perceptual Risk Optimization for Speech Enhancement

1 Jishnu S.
Supervisor: 2Dr. Chandra Sekhar Seelamantula

1Department of Electrical Communication Engineering
2Department of Electrical Engineering
Indian Institute of Science
Bangalore – 560012, India
jishnus@ece.iisc.ernet.in

April 7, 2017
Outline

- Problem statement
- SURE
- Perceptual risk estimation
- Perceptual risk optimization for speech enhancement
- Conclusions
Consider samples of a signal $s_n$, distorted by additive random noise $w_n$. The observation model is given by:

$$x_n = s_n + w_n. \quad n = 1, 2 \cdots$$

Goal: To estimate $s_n$ from $x_n$, by minimizing a suitable distortion metric.
Risk estimation

- **Conventional method**: Obtain an estimate of $s$ by minimizing the distortion function (risk) between estimate $\hat{s} = h(x)$ and $s$,

\[
\hat{s} = \arg \min_{h(x)} E \{ d(h(x), s) \},
\]

where $d$ measure the closeness between $h(x)$ and $s$.

- Direct minimization of cost requires the knowledge of underlying clean signal.

- **Risk Estimation**: Minimize an unbiased estimate of $\mathcal{R}$ to obtain $\hat{s}$.
Basic SURE formulation

- Consider MSE

\[
\mathcal{R} = \mathcal{E} \{ d(h(x), s) \} = \mathcal{E} \left\{ (h(x) - s)^2 \right\} \\
= \mathcal{E} \left\{ s^2 \right\} - 2\mathcal{E} \left\{ h(x) s \right\} + \mathcal{E} \left\{ h(x)^2 \right\}.
\]

where \( x \sim \mathcal{N}(s, \sigma^2) \).

- SURE is an unbiased estimate of MSE obtained using Stein’s lemma. (Stein, 1981)

Let \( Y \) be a real random variable \( \mathcal{N}(0, \sigma^2) \) and let \( h : \mathbb{R} \to \mathbb{R} \) be an indefinite integral of the Lebesgue measurable function \( h' \), essentially the derivative of \( h \). Suppose also that \( \mathcal{E}_Y \{ |h'(Y)| \} < \infty \). Then

\[
\mathcal{E}_Y \{ Yh(Y) \} = \sigma^2 \mathcal{E}_Y \{ h'(Y) \}
\]
SURE

- Using Stein’s lemma: \( \mathcal{E} \{ h(x) s \} = \mathcal{E} \{ h(x) x \} - \sigma^2 \mathcal{E} \{ h'(x) \} \).
- Unbiased estimate of \( \mathcal{R} \) becomes

\[
\hat{\mathcal{R}} = s^2 - 2h(x)x + 2\sigma^2 h'(x) + h(x)^2
\]

i.e. \( \mathcal{R} = \mathcal{E}[\hat{\mathcal{R}}] \). Minimize \( \hat{\mathcal{R}} \) to obtain \( h(x) \).
- Clean speech DCT coefficient estimate, \( h(x_k) = a_k x_k \), where \( a_k \in [0, 1] \) and \( x_k \) is noisy DCT coefficient.
- Optimum pointwise shrinkage parameter \( a_{k,opt} = \arg \min_{a_k} \hat{\mathcal{R}} \)

\[
a_{k,opt} = \left[ 1 - \frac{\sigma^2}{x_k^2} \right]_+ \quad \text{where} \quad [x]_+ = \max(0, x).
\]
Perceptual risk estimation

- Perceptual distortion functions: Itakura-Saito distortion, hyperbolic-cosine (cosh) distortion, weighted cosh distortion, etc. [2].
- Practical noise types are bounded, hence one can model the noise using a truncated Gaussian distribution.
- Assuming observation distribution is truncated gaussian and SNR is high, we propose risk estimate for perceptual distortion functions.
- Minimize perceptual risk estimates to obtain optimum shrinkage estimators.
Itakura Saito (IS) Distortion

- $R_{IS} := \mathcal{E} \left\{ d_{IS}(s_k, \hat{s}_k) \mid |w_k| < |x_k| \right\}$ where

  $$d_{IS}(s_k, \hat{s}_k) = \frac{\hat{s}_k}{s_k} - \log \left( \frac{\hat{s}_k}{s_k} \right) - 1$$

  $$= \frac{\hat{s}_k}{x_k} \left( 1 - \frac{w_k}{x_k} \right)^{-1} - \log (\hat{s}_k) + \log (s_k) - 1$$

  $$= \frac{\hat{s}_k}{x_k} \sum_{n=0}^{\infty} \left( \frac{w_k}{x_k} \right)^n - \log (\hat{s}_k) + \log (s_k) - 1.$$

- Truncating the series beyond $n=4$ using $\hat{s}_k = a_k x_k$ yields

  $$R_{IS} \approx \sum_{n=0}^{4} \mathcal{E} \left( a_k \frac{w_k^n}{x_k^n} \right) - \mathcal{E} \{ \log (a_k x_k) \} + \log (s_k) - 1.$$
Lemma 1

Let $W$ be a real random variable with p.d.f

$$p(w; c_1, c_2, \sigma) = \frac{1}{\sqrt{2\pi\sigma}} \exp \left( -\frac{w^2}{2\sigma^2} \right) \mathbb{1}\{-c_1\sigma < w < c_2\sigma\}$$

where $K = \frac{1}{\sqrt{2\pi\sigma}} \int_{-c_1\sigma}^{c_2\sigma} \exp \left( -\frac{u^2}{2\sigma^2} \right) \, du$ and let $f : \mathbb{R} \to \mathbb{R}$ be an $n$-fold indefinite integral of the Lebesgue measurable function $f^{(n)}$, which is the $n^{th}$ derivative of $f$. Suppose also that $\mathcal{E}\{ |W^{(n-k)}f^{(k)}(W)| \} < \infty$, $c_1\sigma, c_2\sigma >> \sigma$, and $f^{(k)}(W)$ belongs to a class of functions such that $-\sigma^2 f^{(k)}(w) p(w; c_1, c_2, \sigma) \big|_{-c_1\sigma}^{c_2\sigma} \approx 0$, $k = 1, 2, \cdots, n$. Then

$$\mathcal{E}\{ W^n f(W) \} \approx \sigma^2 \mathcal{E}\{ f'(W) W^{n-1} \} + \sigma^2 (n - 1) \mathcal{E}\{ f(W) W^{n-2} \}.$$
Using Lemma 1, the risk $\mathcal{R}_{IS}$ is

$$\mathcal{R}_{IS} = \mathcal{E} \left\{ a_k \left( 1 + 60 \frac{\sigma^6}{x_k^6} + 840 \frac{\sigma^8}{x_k^8} \right) - \log(a_k x_k) \right\} - \log(s_k) - 1.$$  

The unbiased estimate of $\mathcal{R}_{IS}$ is

$$\hat{\mathcal{R}}_{IS} = a_k \left( 1 + 60 \frac{\sigma^6}{x_k^6} + 840 \frac{\sigma^8}{x_k^8} \right) - \log(a_k x_k) - \log(s_k) - 1.$$  

Differentiating $\mathcal{R}_{IS}$ with respect to $a_k$ and equating to zero, we get that

$$a_{k,\text{opt}} = \left[ 1 + 60 \frac{\sigma^6}{\xi^3} + 840 \frac{\sigma^8}{\xi^4} \right]^{-1}$$

where $\xi_k = \frac{x_k^2}{\sigma^2}$.  

Table: Optimal shrinkage parameters for different perceptual risk estimates.

<table>
<thead>
<tr>
<th>Risk</th>
<th>$d(s_k, \hat{s}_k)$</th>
<th>$a_{opt}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>log MSE</td>
<td>$(\log \frac{\hat{s}_k}{s_k})^2$</td>
<td>$\exp \left( \frac{0.5}{\xi_k} - \frac{0.75}{\xi_k^2} - \frac{10}{\xi_k^3} - \frac{210}{\xi_k^4} \right)$</td>
</tr>
<tr>
<td>WE</td>
<td>$\frac{(\hat{s}_k - s_k)^2}{s_k}$</td>
<td>$\left[ 1 + \frac{1}{\xi_k} - \frac{1}{\xi_k^2} + \frac{48}{\xi_k^3} + \frac{360}{\xi_k^4} \right]^{-1}$</td>
</tr>
<tr>
<td>IS-II</td>
<td>$\frac{\hat{s}_k^2}{s_k^2} - \log \frac{\hat{s}_k^2}{s_k^2} - 1$</td>
<td>$\left[ 1 - \frac{1}{\xi_k} + \frac{24}{\xi_k^2} + \frac{360}{\xi_k^3} + \frac{4200}{\xi_k^4} \right]^{-\frac{1}{2}}$</td>
</tr>
<tr>
<td>COSH</td>
<td>$\frac{1}{2} \left[ \frac{s_k}{\hat{s}_k} + \frac{\hat{s}_k}{s_k} \right] - 1$</td>
<td>$\sqrt{1 + \frac{1}{\xi_k}} / \sqrt{1 + 60 \frac{1}{\xi_k^3} + 840 \frac{1}{\xi_k^4}}$</td>
</tr>
<tr>
<td>WCOSH</td>
<td>$\left[ \frac{s_k}{\hat{s}_k} + \frac{\hat{s}_k}{s_k} - 1 \right] \frac{1}{s_k^p}$</td>
<td>$\left[ 1 - \frac{1}{\xi_k} + \frac{3}{\xi_k^2} + \frac{420}{\xi_k^3} + \frac{9450}{\xi_k^4} \right]^{-\frac{1}{2}}$</td>
</tr>
</tbody>
</table>

where $\xi_k = \frac{\chi_k^2}{\sigma^2}$.
Perceptual Risk Optimization for Speech Enhancement

Performance Evaluation

Figure: Performance comparison of different denoising algorithms.
Conclusion

- Introduced the notion of risk estimation for single-channel speech enhancement.
- We proposed risk estimates for perceptual distortion metrics and minimize to obtain the optimum denoising function.
- For SNR greater than 5 dB, the proposed approach resulted in better denoising performance than the benchmarking techniques.


THANK YOU