2014 IEEE International Conference on Acoustic, Speech and Signal Processing (ICASSP)

# ML ESTIMATION OF MEMORYLESS NONLINEAR DISTORTIONS IN AUDIO SIGNALS

*Flávio R. Ávila*

DETEL/FEN
Rio de Janeiro State University
20550-900 Rio de Janeiro, Brazil
flavio.avila@uerj.br

*Luiz W. P. Biscainho*

DEL/Poli & PEE/COPPE
Federal University of Rio de Janeiro
21941-972 Rio de Janeiro, Brazil
wagner@smt.ufrj.br

**ABSTRACT**

Many real-world signals are subjected to nonlinear distortions that can be approximately modeled as memoryless and invertible. In Audio applications, they are typical of magnetic recordings but can also result of dynamic compression employed in vinyl recordings etc. Such an effect can be disturbing to a modern audience which is used to higher quality material. This paper proposes an iterative algorithm to maximize the likelihood function of the distortion function parameters, based solely on samples of the degraded signal, and then recover the original signal. The method assumes the original signal to be autoregressive and Gaussian in short sections—a standard model for audio—and the nonlinearity to be time-invariant throughout the signal, thus allowing the use of all samples in the model estimation. Additionally, a simple and time-efficient alternative technique to estimate the nonlinear function is proposed; it can be used either as a fast and reliable stand-alone procedure or as a initialization routine for the more sophisticated maximum likelihood approach. The robustness of the proposed techniques is verified through application to artificial and real signals nonlinearly distorted.

***Index Terms*—** Nonlinear; Restoration; Maximum Likelihood; Gauss-Newton

## 1. INTRODUCTION

The restoration of nonlinearly distorted audio material is particularly challenging because of the generality of this kind of defect. Successful attempts to address the reduction of memoryless nonlinearities have been described for special contexts, such as magnetic recorded movie soundtracks [1] and horn loudspeakers [2], where simpler models are suitable.

This paper proposes a model-based statistical method to identify a memoryless nonlinear distortion whose inversion enables one to recover the original signal. The method assumes the original signal to be piece-wise auto-regressive with white Gaussian excitation, and the nonlinear curve to be parameterizable by a polynomial expansion, which can model a wide range of distortions. An iterative algorithm based on the Gauss-Newton method is implemented to find the Maximum Likelihood (ML) of the unknown parameters. In addition, an alternative simple and robust scheme to estimate the parameters, which can also be used as a starting point for the iterative algorithm, is described. By assuming the distortion function equals the identity around the origin, and the signal to be Gaussian in short sections, this algorithm first estimates the original signal variance, and then calculates the nonlinear curve that would make the observed distribution of the degraded signal maximally consistent with the expected distribution of the original signal.

After this Inroduction, Section 2 states the parametric model adopted for both the audio signal and the nonlinear curve. In Section 3, an identification technique that exploits the expected distribution of audio signals is described. Section 4 presents the calculation of the likelihood function for the unknown parameters given the degraded audio samples, and the corresponding ML solution using the Gauss-Newton algorithm. In Section 5, the application of both algorithms to some nonlinearly distorted signals shows their efficacy. Finally, conclusions are drawn in Section 6.

Thanks to CAPES, CNPq, and FAPERJ Brazilian agencies for funding.

## 2. MODELING THE NONLINEAR DISTORTION

A system is said to be memoryless when its output at a given instant depends only on its input at that instant. In audio, a usual memoryless nonlinear distortion has the form of a *soft clipping*, illustrated in Fig. 1. The original signal $x_n$ is assumed to be generated by an all-pole filter $A(z)$, supposed constant in short time intervals, and excited by white Gaus-

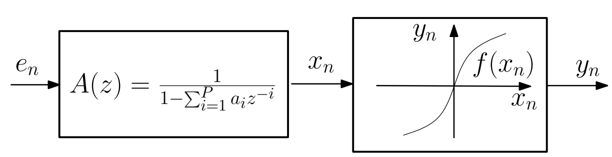


**Fig. 1**. Audio signal corrupted by memoryless nonlinearity.

sian noise $e_n$—a model successfully applied to audio signals in the restoration context [3]. The degradation is imposed on the signal by a time-invariant and memoryless nonlinear function $f(.)$ assumed to be anti-symmetric, monotonically increasing (thus invertible), and close to the identity around the origin. Time-invariance assumption is justifiable if one considers the behavior of the recording or reproduction system that imposes the distortion to be fixed throughout time.

Since this model disregards noise (considered as negligible), the problem reduces to the estimation of the curve $f(.)$, or rather its inverse $g(.)$, from which the original signal can be readily obtained. As in [4, 5, 6], this function is modeled as a polynomial expansion with unknown parameters $m_i$:

$$x_n = y_n + m_1 {y_n}^3 + m_2 {y_n}^5 + \ldots + m_M {y_n}^{2M+1}, \quad (1)$$

where the even coefficients are null, so that $g(y) = -g(-y)$, and the linear term coefficient equals 1, so that $g'(0) = 1$. The model order $M$ is unknown, but chosen to be sufficiently high so as to guarantee a good approximation of the curve. Eq (1) can be written in vectorial form as $\mathbf{x} = \mathbf{y} + \mathbf{Y}\mathbf{m}$, where $\mathbf{m} = [m_1, \ldots, m_M]^T$, the element $Y_{ij}$ of $\mathbf{Y}$ is given by $Y_{ij} = y_i^{2j+1}$ with $i \in \{1, \ldots, N\}$ and $j \in \{1, \ldots, M\}$, and $N$ is the number of samples in the degraded signal.

Audio signals can be regarded as stationary for short-sections (typically 20-ms), during which both the statistics of $e_n$ and the coefficients $\mathbf{a} = \{a_1, \ldots, a_P\}$ of its AR model can be held fixed. In addition, for sufficiently high order $P$, the excitation $e_n$ can be taken as composed by i.i.d. zero-mean Gaussian samples with unknown variance. Many practical techniques segment the audio signal into fixed-length blocks which are separately processed, thus taking advantage of the stationarity assumption. The same approach is adopted here: the signal is divided into $B$ $L$-length blocks. In a given block $j$, signal samples are denoted as $\mathbf{x}^{(j)}$, AR parameters as $\mathbf{a}^{(j)}$, and the excitation variance as $\sigma_e^{2^{(j)}}$. The coefficient-vector $\mathbf{m}$ of the nonlinear curve is assumed constant throughout the signal, thus affecting all samples the same way.

## 3. IDENTIFICATION BASED ON DISTRIBUTION EQUALIZATION

In [7], a memoryless nonlinear distortion is identified by comparison between the histogram of the degraded signal and an expected histogram of the original signal, estimated from a representative database of undistorted signals. Although inspired by that overall idea, the more general approach proposed here does not require such a database and thus can be applied to a larger variety of signal categories.

The key idea of the method is to explore the expected shape of the sample histogram computed in short sections of a typical undistorted signal. The observation of a large set of 1000-sample blocks of undistorted audio signals indicates that the signal kurtosis clusters around 3, which suggests they can be loosely seem as Gaussian. Even if this hypothesis is not always accurate, the combination of estimates performed over many sections tends to cancel out the individual errors and yield perceptually good results.

By assuming each block $j$ of the original signal is Gaussian with unknown variance $\sigma_x^{2^{(j)}}$, the proposed algorithm searches for the nonlinear transformation that restores Gaussianity from the observed non-Gaussian degraded signal. Afterwards, it combines all blocks in order to produce a global estimate of the nonlinear function parameters. A key assumption is the linearity in the vicinity of the origin, which allows one to estimate the variance of the original signal.

Since the transformation is memoryless and invertible:

$$F_{Y^{(j)}}(y) = F_{X^{(j)}}(g(y)), \quad (2)$$

where $F_{Y^{(j)}}(.)$ and $F_{X^{(j)}}(.)$ are the cumulative distributions of $\mathbf{y}$ and $\mathbf{x}$ in a certain block $j$.

By applying $F^{-1}_{X^{(j)}}(.)$ to both sides of the equation, $g(y)$ can be calculated as:

$$g(y) = F^{-1}_{X^{(j)}}(F_{Y^{(j)}}(y)). \quad (3)$$

The cumulative distribution $F_{Y^{(j)}}(y)$ can be approximated by the empirical distribution [8]:

$$F_{Y^{(j)}}(y) \approx \hat{F}_{Y^{(j)}}(y) = \frac{1}{N}\#\{n, y_n < y\}. \quad (4)$$

By replacing this estimate into (3) and taking into account the Gaussian assumption for $X$, $g(y)$ can be estimated as

$$\hat{g}(y) = \Phi^{-1}_{\sigma_x^{2^{(j)}}}\left(\hat{F}_{Y^{(j)}}(y)\right), \quad (5)$$

where $\Phi_{\sigma_x^{2^{(j)}}}(x)$ is the cumulative distribution of a zero-mean Gaussian variable with variance $\sigma_x^{2^{(j)}}$. This variance can be estimated by using the assumption that $g'(0) = 1$. By differentiating both sides of (2) with respect to $y$, one finds that $f_{Y^{(j)}}(0) = f_{X^{(j)}}(0) = \frac{1}{\sqrt{2\pi\sigma_x^2}}$, i.e.

$$\sigma_x^{2^{(j)}} = \frac{1}{2\pi f_{Y^{(j)}}(0)^2}. \quad (6)$$

Thus, in order to estimate $\sigma_x^{2^{(j)}}$, one just needs a good estimate for $f_{Y^{(j)}}(0)$—which can be easily obtained by calculating the fraction of samples lying in a small interval around the origin and dividing it by the length of this interval.

### 3.1. Coefficient estimation

A simple way to estimate the model coefficients from its non-parametric estimation given in (5) is by applying the Least-Squares method, according to the following procedure:

1. For every $n \in \{1, \ldots, N\}$, form the pair $(y_n, \hat{g}(y_n))$;
2. Build matrix $\mathbf{Y}$ as defined in Section 2;
3. Estimate the parameter vector $\mathbf{m}$ as

$$\hat{\mathbf{m}} = (\mathbf{Y}^T\mathbf{Y})^{-1}\mathbf{Y}^T(\hat{g}(\mathbf{y}) - \mathbf{y}). \quad (7)$$

## 4. IDENTIFICATION BASED ON LIKELIHOOD MAXIMIZATION

Let $\mathbf{a} = \{\mathbf{a}^{(1)}, \ldots, \mathbf{a}^{(B)}\}$ and $\sigma_e^2 = \{\sigma_e^{2^{(1)}}, \ldots, \sigma_e^{2^{(B)}}\}$. The likelihood function is defined as $l(\mathbf{a},\sigma_e^2,\mathbf{m};\mathbf{y}) = p(\mathbf{y}|\mathbf{a},\sigma_e^2,\mathbf{m})$. After successively applying the Bayes' rule and using the formula for random variable transformation as well as the Gaussian assumption for $e_n$,

$$l(\mathbf{a}, \sigma_e^2, \mathbf{m}; \mathbf{y}) = \left( \prod_{n=P+1}^{N} g'(y_n) \right) \times \prod_{j=1}^{B} \left( \frac{1}{\sqrt{2\pi\sigma_e^{2^{(j)}}}} \right)^L \exp\left\{ -\frac{1}{2\sigma_e^{2^{(j)}}} \mathbf{e}^{(j)^T} \mathbf{e}^{(j)} \right\},$$

where $\mathbf{e}^{(j)} = \{e_{(j-1)L+P+1}, \ldots, e_{jL+P}\} = \mathbf{e}^{(j)} = \mathbf{A}^{(j)}\mathbf{x}^{(j)}$, with $\mathbf{A}^{(j)}$ and $\mathbf{x}^{(j)}$ defined in such a way that $e_n = x_n - \sum_{i=1}^{P} a_i x_{(n-i)}$. Each element of $\mathbf{x}^{(j)}$ can be calculated from $x_n = g(y_n)$.

### 4.1. Maximum Likelihood Estimation

This approach is inspired by [5], which tackles the problem of blind identification of Wiener systems with uncorrelated Gaussian input. The problem addressed here is more general, since both the input and the linear part of the model are time-varying.

First, the negative log-likelihood is computed:

$$L(\mathbf{a}, \mathbf{m}, \sigma_e^2) = -\log\left[ l(\mathbf{a}, \sigma_e^2, \mathbf{m}; \mathbf{y}) \right] = \frac{BL}{2}\log(2\pi) + \frac{L}{2}\sum_{j=1}^{B} \log(\sigma_e^{2^{(j)}}) - \sum_{n=P+1}^{N} \log\{g'(y_n)\} + \sum_{j=1}^{B} \left( \frac{1}{2\sigma_e^{2^{(j)}}} \mathbf{e}^{(j)^T}\mathbf{e}^{(j)} \right). \quad (8)$$

By equating to zero the derivative of $L(\mathbf{a}, \mathbf{m}, \sigma_e^2)$ with respect to each $\sigma_e^{2^{(j)}}$, one obtains

$$\sigma_e^{2^{(j)}} = \frac{\mathbf{e}^{(j)^T}\mathbf{e}^{(j)}}{L}. \quad (9)$$

By replacing $\sigma_e^{2^{(j)}}$ into (8) and following the same procedure presented in [5], it can be shown that minimizing $L(\mathbf{a}, \mathbf{m}, \sigma_e^2)$ is equivalent to minimizing the following objective function:

$$V(\mathbf{a}, \mathbf{m}) = \sum_{n=P+1}^{N} r_n^2, \quad (10)$$

where

$$r_n = g(\mathbf{m}) e_n(\mathbf{a}, \mathbf{m}), \quad (11)$$

$$g(\mathbf{m}) = \exp\left\{ -\frac{1}{N-P} \sum_{n=P+1}^{N} \log\{g'(y_n)\} \right\}, \quad (12)$$

$$e_n(\mathbf{a}, \mathbf{m}) = g(y_n) - \sum_{i=1}^{P} a_i^{(\lceil (n-P)/L \rceil)} g(y_{n-i}). \quad (13)$$

A suitable strategy to perform this minimization (and thus maximize the likelihood) is the Gauss-Newton method, a simplification of the Newton method applicable when the objective function can be written as a sum of squares (as in this case), summarized as follows. Let

$$F : \mathbb{R}^D \to \mathbb{R}, F(\theta) = \sum_{n=1}^{N} r_n^2(\boldsymbol{\theta}) \quad (14)$$

be the objective function.

Starting from an estimate $\boldsymbol{\theta}_0$, one can reach the optimum solution by successively updating the parameters according to $\boldsymbol{\theta}^{(k+1)} = \boldsymbol{\theta}^{(k)} + \boldsymbol{\Delta}^{(k)}$, where

$$\boldsymbol{\Delta}^{(k)} = -\left( \mathbf{J}(\boldsymbol{\theta}^{(k)})^T \mathbf{J}(\boldsymbol{\theta}^{(k)}) \right)^{-1} \mathbf{J}(\boldsymbol{\theta}^{(k)})^T \mathbf{r}(\boldsymbol{\theta}^{(k)}). \quad (15)$$

Each element of the Jacobian $\mathbf{J}(\boldsymbol{\theta})$ is defined as:

$$\mathbf{J}(\boldsymbol{\theta})_{ij} = \frac{\partial r_i(\boldsymbol{\theta})}{\partial \theta_j}, i = \{1, 2, \ldots, N\}, j = \{1, 2, \ldots, D\}, \quad (16)$$

and the elements of vector $\mathbf{r}(\boldsymbol{\theta})$ are $r_n(\boldsymbol{\theta})$, $n = \{1, 2, \ldots, N\}$.

For the problem at hand, $\boldsymbol{\theta} = \{\mathbf{m}, \mathbf{a}^{(1)}, \mathbf{a}^{(2)}, \ldots, \mathbf{a}^{(B)}\}$. The derivatives of $r_n(\boldsymbol{\theta})$ with respect to each element of $\boldsymbol{\theta}$ can be easily computed from (11) and will be omitted here.

## 5. RESULTS

To validate the proposed methods, first an artificially generated signal from an AR model of order $P = 6$ is distorted with a curve defined by coefficients $m_1 = 5$ and $m_2 = 30$. Both the method of distribution equalization (Sec. 3) and the ML are applied to the degraded signal, yielding the results shown in Fig. 2. The scatter plot allows us to see how much nonlinear distortion is still present in the restored signals. While both methods reduce significantly the nonlinearity, the ML approach yields better results: the restored signal is expected to almost coincide with the original one.

Fig. 3 shows that the Gauss-Newton algorithm converges quickly to the maximum likelihood estimates, which are very close to the correct parameters (represented as red squares in the plot). Fig. 4 shows histograms of the estimated parameters, obtained upon 1000 applications of the method on data generated according to the bootstrap principle [8]. This figure indicates that the confidence interval is indeed very narrow, which is a consequence of the large amount of data being used as compared to the number of parameters.

In order to assess the methods in a more realistic scenario, we generated a database of audio signals comprised of excerpts from a comprehensive set of classical music tracks artificially degraded by non-polynomial distortions (which

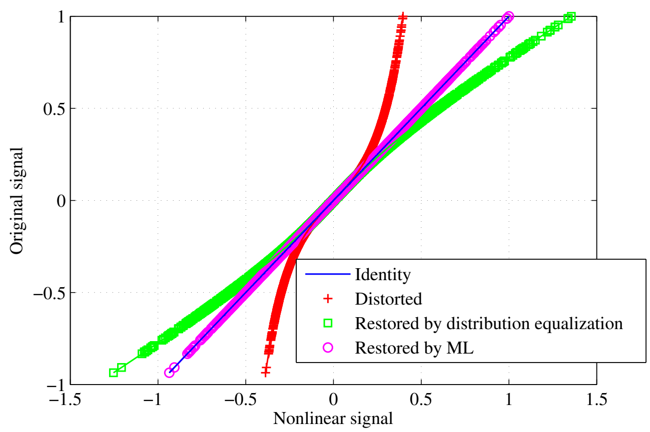


**Fig. 2**. Comparison of algorithm performances for an artificial signal.

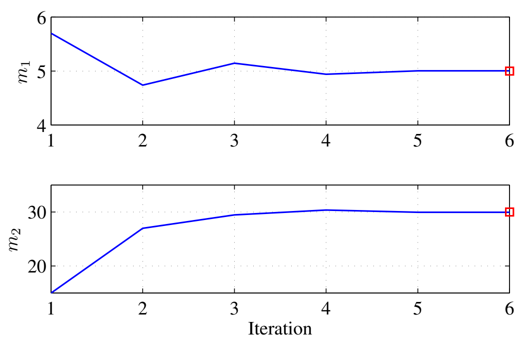


**Fig. 3**. Evolution of Taylor expansion coefficients generated by the ML method.

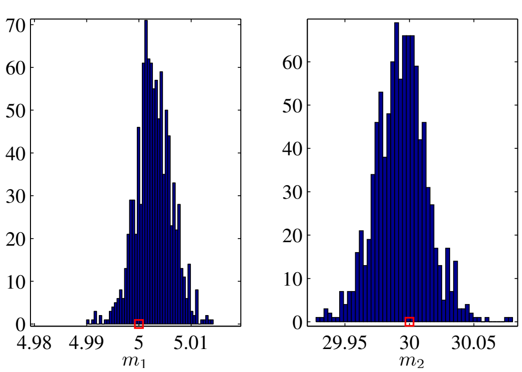


**Fig. 4**. Distribution of estimates for Taylor expansion coefficients, generated with the bootstrap method. The red square indicates the correct value.

makes the curve fitting harder for the polynomial model). For instance, Fig. 5 shows the results corresponding to a given audio signal distorted by $y_n = \arctan(x_n)$. One can see that both algorithms are able to reduce the signal nonlinearity, but the ML solution produces significantly better results. To allow independent evaluation of results, examples with music signals are provided in [9].

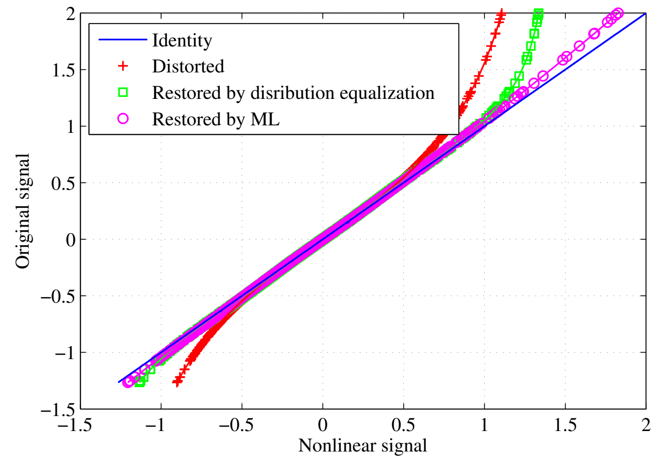


**Fig. 5**. Comparison of algorithm performances for a real signal.

## 6. CONCLUSIONS

This paper dealt with the restoration of audio signals that have been distorted by memoryless nonlinearity. Two techniques to identify the nonlinear distortion function, and thus reconstruct the original signal, were presented. Another potential application of the methods is on digital modeling of tube amplifiers, which can be partially characterized by a nonlinear distortion [10].

The assumption of signal Gaussianity proved to be instrumental in finding the degradation curve. In a model-based framework, the autoregressive model with Gaussian excitation was shown to yield better results than the Gaussian assumption alone. In the latter case, the Taylor series expansion of the nonlinear curve is effective in describing a wide range of memoryless nonlinearities and leads to a simple maximum likelihood estimate of the curve, which can be implemented via the Gauss-Newton algorithm.

Compared to previous works in the literature, the proposed methods have the advantage of being robust to a variety of nonlinearity formats, as long as they are memoryless and invertible. The described approach also establishes a convenient framework to build more sophisticated models to address e.g. additive noise and nonlinearity with memory.

## 7. REFERENCES


[1] T. B. Bakó and T. Dabóczi, “Restoration of nonlinearly distorted audio with the application to old motion pictures,” in *Proceedings of the AES 20th International Conference*, Budapest, Hungary, October 2001, AES, pp. 191–198.

[2] W. Klippel, “Compensation for nonlinear distortion of horn loudspeakers by digital signal processing,” *Journal of the Audio Engineering Society*, vol. 44, no. 11, pp. 964–972, November 1996.

[3] S. J. Godsill and P. J. W. Rayner, *Digital Audio Restoration - A Statistical Model Based Approach*, Cambridge University Press, Cambridge, UK, 1998.

[4] S. J. Godsill, P. J. W. Rayner, and O. Cappé, “Digital audio restoration,” in *Applications of Digital Signal Processing to Audio and Acoustics*, M. Kahrs and K. Brandenburg, Eds., chapter 4, pp. 133–193. Kluwer, New York, USA, 1998.

[5] L. Vanbeylen, R. Pintelon, and J. Schoukens, “Blind maximum-likelihood identification of Wiener systems,” *IEEE Transactions on Signal Processing*, vol. 57, no. 9, pp. 3017–3029, August 2009.

[6] G. Picard and O. Cappé, “Blind identification of Hammerstein nonlinear distortion models,” in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, USA, October 2003, IEEE, pp. 17–20.

[7] S. A. White, “Restoration of nonlinearly distorted audio by histogram equalization,” *Journal of the Audio Engineering Society*, vol. 30, no. 11, pp. 828–832, November 1982.

[8] L. Wasserman, *All of Statistics*, Springer, New York, USA, 2010.

[9] F. Ávila, “Examples of Restored Signals,” 2013, http://www.prosaico.uerj.br/professores/flavio/pessoal.html/.

[10] J. Pakarinen and D. T. Yeh, “A review of digital techniques for modeling vacuum-tube guitar amplifiers,” *Computer Music Journal*, vol. 33, no. 2, pp. 85–100, – 2009.