Improved-Speed Parameter Tuning of Deconvolution Algorithm

Tamás B. Bakó and Tamás Dabóczi

Abstract

Deconvolution (or inverse filtering) is applied if a time, space or other domain signal is distorted by a measurement system, and one wants to reconstruct the signal to be measured. This is an estimation process since the measurement is always corrupted by noise. The estimation problem is usually ill-posed, meaning that even very slight changes in the measured signal caused by a moderate noise will cause extreme fluctuation in the estimate. Different kind of regularizations are applied to suppress this noise amplification which can be accomplished only at the price of bias in the estimate.

The big challenge is to set the trade-off between the variance and bias. Most deconvolution algorithms can be tuned with one parameter or several parameters. This paper presents a systematic algorithm to optimize the parameters of deconvolution algorithms. Its advantage is that it is directly derived from the input error criterion with some approximations and spectral modeling. Spectral models of the signals are built automatically, thus no user interaction is needed, the procedure is completely automatic. The optimization method is based on a former algorithm of one of the authors, for which a significant improvement in computational speed (apprx. 100..150 times) has been achieved for Tikhonov type deconvolution method by replacing one of the iteration processes by a closed form expression. The convergence of the method is analyzed. The usefulness of the proposed method is demonstrated on simulated and measured data. Error benchmarks and robustness comparisons with competitor algorithms are also presented.

Index Terms

regularization, deconvolution, inverse filtering, Tikhonov's regularization, optimization

I. Introduction

Measurement systems can be characterized by different quality parameters. We focus now on systems which aim at measuring time-, space- or other domain signals, and the main limitation is the finite bandwidth of the measuring device (i.e. we assume a linear and time-shift invariant model for the measurement system). In this case the acquired waveform from the measurement system is a distorted version of the true one. On top of that the measurement is always contaminated by additive noise and/or interference. In such cases one wants to reconstruct the distortion by means of digital post-processing of the measured signal (deconvolution or inverse filtering). This procedure is usually ill-posed, which means that the estimated signal has a very large variance due to even a small amount of noise. To overcome on this problem signal reconstruction is combined with some kind of noise suppression, which however leads to distortion in the estimate (Fig. 1). The big challenge is to find the optimal trade-off between the large variance and large bias. There were many attempts to suppress the amplified noise during the inverse filtering process. Our study focuses on methods which tune the trade-off by a limited number of parameters (we will call them parametric inverse filtering methods).

The paper is organized as follows. Section II briefly summarizes the most common inverse filtering methods. Section III collects previous attempts to tune the parameters of inverse filtering methods.

T. B. Bakó works at Morgan Stanley Hungary Analytics Ltd. e-mail: bako@mit.bme.hu

T. Dabóczi is with the Department of Measurement and Information Systems, Budapest University of Technology and Economics, 1521 Budapest, Hungary. e-mail: daboczi@mit.bme.hu

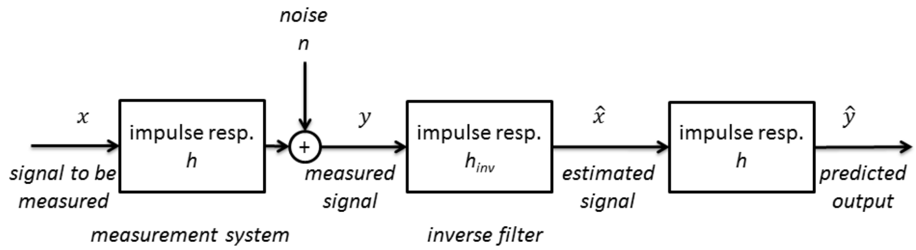

Fig. 1. Block-scheme of the investigated ill-posed problem.

Our model based parameter optimization is introduced in Section IV. In Section V, simplification on the method will be proposed, which significantly increases calculation speed. Convergence property of the new algorithm is analyzed. Next the sensitivity of the algorithm on noise level estimate is derived. Section VII shows the usefulness of the method on simulated and measured data, together with comparison of the performance of other methods. The novelty of the paper is the improvement on the speed and the convergence analysis.

II. Inverse filtering methods

An early approach to solve ill-conditioned problems is Tikhonov's theory [1]. He introduced regularization operators to reformulate the original problem and to control the behavior of the estimate. The regularization operators incorporate our a priori knowledge about the signal to be reconstructed. In practice mostly one, rarely a few operators are used ([2–5]). Typical regularization operators are the energy-, smoothness- and higher order derivatives of the estimated signal.

A simple method to suppress noise during inverse filtering is to apply a lowpass filter before compensation of the distortion of measurement system [6]. The cutoff frequency of the filter is the parameter to control the trade-off between noisy and biased estimates. It can be applied only for systems having lowpass nature, and signals concentrating their energy to the lower frequency bands. If we have a priori knowledge about the signal to be reconstructed, a matched-or cross-correlation filter can enhance the signal-to-noise ratio (power of useful signal compared to that of the noise) on a more effective way [7].

Iterative inverse filtering processes have the advantage that amplitude constraints are easy to incorporate [8–10]. However, the computational demand is very large. Noise suppression is controlled by the number of iterations.

Stochastic approaches can take not only output noise, but also process noise into account [13]. Optimal filter can be derived only with some assumptions about the knowledge of the signal to be measured or at least about some parameters of it (e.g. authors of [14] assume autoregressive model for the signal).

As different algorithms utilize different properties of the signal to be reconstructed (amplitude constraint, lowpass nature, smoothness etc.), and different level of a priori knowledge, the performance of inverse filtering depends both on the signal and on deconvolution method.

III. Optimization of parametric inverse filtering methods

The field of inverse filtering algorithms is more or less thoroughly investigated. A much less developed field is the optimization of the algorithms, i.e. tuning the parameters of the inverse filter to receive the optimal reconstruction from the point of view of an error norm.

One of the early attempts to optimize the frequency domain version of Tikhonov's regularization is Nahman's work [4, 5]. He observed specific behavior of the rounding error of computation as a function of regularization parameter and created a frequency domain regularization technique. However, the idea is heuristic, and strongly depends on the implementation of the Fourier-transform and on the processor architecture.

Parruck et al. derived a method for non-parametric system identification, which is also a deconvolution problem [10]. The estimated impulse response is integrated to receive an estimate of the step-response. The tail part of the step-like waveform is further investigated and the optimum is set where the mean value and the standard deviation satisfy certain conditions. Unfortunately, manual interaction is still required as the solution is not unique.

A similar method was developed by Bertocco et al. to reconstruct step-like waveforms [2]. The idea is that the noise on the measurement should have uniform distribution in the whole record length. The method is easy to implement, and showed a good performance, however, the set of signals is limited to step-like waveforms.

In Morozov's algorithm [11], the idea is to utilize a priori information about the level of noise on the measurement. The optimal solution is defined for which the output error (difference between the measured signal and predicted output) is very close to the a priori noise level. The drawback of the algorithm is that it requires an accurate measurement of the noise level.

Younan et al. investigate also the output error, however, not the predicted output, but the system output itself. They separate inverse filtering to a lowpass filtering (truncation of FFT) and a successive reconstruction step [6]. The optimality is defined for which the randomness of the truncated part of the signal is maximal, which is measured by serial correlation coefficient. The drawback of the method is that the regularization method is restricted to sharp lowpass filtering, consolidated by windowing.

Dhaene et al. developed a method to handle two-parameter inverse filtering (Tikhonov's regularization with energy and smoothness constraints) [15]. Several noise factors need to satisfy certain conditions which need to be evaluated on three-dimensional or contour plots. Although handling two parameters provides more flexibility from the point of view of reconstruction performance, the procedure requires manual separation of the transfer function into pass-, attenuation- and stop-bands.

A classical method for parameter optimization of regularization problems is the so called "L-curve" method. Here the norm of the regularized estimate (estimated signal after applying the regularization operator) is plotted as a function of the norm of the output error. This curve has a monotonic decay with a characteristic corner, which is claimed to correspond to the optimal regularization value. The most well known algorithm to select this characteristic point automatically is from Hansen to find the maximum curvature [16], [17]. Even if the L-curve idea (output error equals regularized estimate) seems to be reasonable, it is not proven that it is optimal. Despite this limitation, L-curve method is popular in many applications (e.g. [18]).

Roy and Souders also developed a heuristic algorithm which can weight the error in the time domain, allowing to emphasize important parts of reconstruction [19]. The performance of the algorithm is very good for step-like waveforms.

Olofson dealt with ultrasonic signals, and developed a technique to compensate for the effect of clipping, if the AD converter is slightly overloaded in amplitude [20]. In this method the distribution of the signal to be reconstructed needs to be known. Many times we miss this information.

Szolgay and Sziranyi determined optimal stopping criterion for image deblurring, assuming Richardson-Lucy iterative restoration method [21]. This method does not use any regularization operator. They showed promising results for image reconstruction, but in general case of time-domain deterministic signals, regularization operators are required to suppress the amplified noise.

Kido et al. developed automated image restoration technique for X-ray photography. Although the procedure is automatic at the end, the parameters are empirically set, i.e. the optimal parameter set is image type specific [22]. Similarly, there are many automated algorithms for image restoration, but they rely on some specificity of the fact, that the two dimensional signal is an image of some real object (e.g. [23]), or require the knowledge of the distribution of the unknown measured data [24].

IV. Model based parameter optimization

In this section we will introduce our model based parameter optimization method, which can automatically set the parameter(s) of (parametric) inverse filtering methods. Next section will show the suggested modification, which improves the computation speed.

The block-scheme of the ill-conditioned problem can be seen in Fig. 1. The signal to be measured (the signal of interest) is distorted by a linear filter, h, then it is contaminated by noise. The reconstruction implies pushing the observed signal, y through an "inverse filter", h_{inv}, in order to get a proper estimate of x. This estimate is denoted by $\hat{x}$. If we push $\hat{x}$ through the (simulated) copy of the original linear filter, we get an estimate about the observed signal (predicted output). This is denoted by $\hat{y}$. In the frequency domain convolution operation reduces to multiplication:

$$Y(f) = H(f)X(f) + N(f) \tag{1}$$

$$\hat{X}(f) = H_{inv}(f)Y(f) = H(f)H_{inv}(f)X(f) + H_{inv}(f)N(f) \tag{2}$$

$$\hat{Y}(f) = H(f)\hat{X}(f) \tag{3}$$

where capital letters correspond to the Fourier transform (in our case DFT) of the signals. Without any noise reduction or regularization the inverse filter is the reciprocal of the transfer function of the system:

$$H_{inv}(f) = \frac{1}{H(f)} \qquad \hat{X}(f) = X(f) + \frac{N(f)}{H(f)} \tag{4}$$

At those regions, where the input signal is strongly damped by $H(f)$, performing the inverse filtering may amplify the added noise so much that the result will be useless. Therefore, we are looking for an "optimal inverse filter", which suppresses the noise but keeps the bias also on a limited level. We define the optimum as the minimum of some kind of norm of the difference of x and $\hat{x}$. Please note that this is the input error, and not the usual output error, which minimization leads to large noise amplification.

As it is mentioned in [25, 26] input error criterion cannot be directly used because the input signal (or its spectrum) is unknown. However, our aim is still to define the optimal solution as the one having the minimal input error, even if we cannot perfectly calculate it. There are alternative criteria (cited in the previous section) which minimize alternative function (as e.g. output error, some characteristic points of curves etc.), but their aim is also to have a solution close to the one of the input error criterion. Our approach has the advantage that it is directly derived from the input error criterion, with some spectral approximations. This includes a spectral modeling of signals and some approximations of the input error expression in the frequency domain. Our method requires approximate spectral models of the noise and input signal. Only the magnitude of the spectrum is required, the phase information is not utilized. The spectral models are automatically extracted from the measured data, thus no user interaction is required, the procedure is fully automatic [3, 27, 28].

By using Euclidean norm, the aim is to minimize

$$\sum_t (x(t) - \hat{x}(t))^2 \tag{5}$$

Applying Parseval's theorem the input error criterion can be written in the frequency domain as

$$\frac{1}{N_f} \sum_f \left(X(f) - \hat{X}(f) \right) \left(\overline{X}(f) - \overline{\hat{X}}(f) \right) \tag{6}$$

Substituting equations (1), (2) and (3) into (6), and omitting the normalization by N_f:

$$\begin{aligned}
&\sum_f X(f)\overline{X}(f) \left(1 - H(f)H_{inv}(f)\right) \left(1 - \overline{H}(f)\overline{H}_{inv}(f)\right) \\
&+\sum_f N(f)\overline{N}(f) H_{inv}(f)\overline{H}_{inv}(f) \\
&-\sum_f X(f)\overline{N}(f) \left(1 - H(f)H_{inv}(f)\right) \overline{H}_{inv}(f) \\
&-\sum_f \overline{X}(f)N(f) \left(1 - \overline{H}(f)\overline{H}_{inv}(f)\right) H_{inv}(f)
\end{aligned} \tag{7}$$

It has been proven that the cross terms of signal and noise can be neglected and it is enough to minimize only the rest (see e.g. [3]):

$$\begin{aligned}
&\sum_f X(f)\overline{X}(f) \left(1 - H(f)H_{inv}(f)\right) \left(1 - \overline{H}(f)\overline{H}_{inv}(f)\right) \\
&\quad +\sum_f N(f)\overline{N}(f) H_{inv}(f)\overline{H}_{inv}(f) \\
&= \sum_f |X(f)|^2 \left(1 - H(f)H_{inv}(f)\right) \left(1 - \overline{H}(f)\overline{H}_{inv}(f)\right) \\
&\quad +\sum_f |N(f)|^2 H_{inv}(f)\overline{H}_{inv}(f)
\end{aligned} \tag{8}$$

This step was the approximation of the input error criterion by neglecting a term. The next approximation is the substitution of the absolute value of signal spectra by spectral models. The minimization can be accomplished by an iterative method, where we start from rough spectral models.

$$\begin{aligned}
cost = &\sum_f |X_m(f)|^2 \left(1 - H(f)H_{inv}(f)\right) \left(1 - \overline{H}(f)\overline{H}_{inv}(f)\right) \\
&+\sum_f |N_m(f)|^2 H_{inv}(f)\overline{H}_{inv}(f)
\end{aligned} \tag{9}$$

where *cost* denotes the cost function to be minimized, $|X_m(f)|$ and $|N_m(f)|$ mean spectral models for the absolute values. The absolute value of the noise spectrum is assumed to be constant through the whole frequency range, and its height is estimated based on the signal energy at the stopband region. Spectral model of the signal to be measured is extracted from the measurement (absolute value of the spectrum of the observation) as an initial guess, and refined with an interactive process. Using the initial spectral models the optimal parameter set can be derived by minimizing (9), leading to our first estimate of $H_{inv}(f)$. Now, we can refine our spectral model of the input signal by the absolute value of the spectrum of reconstructed signal. (9) is minimized again to acquire a better estimate of the input signal, which provides again a new spectral model. This procedure is iteratively repeated, until the change in the parameters become small enough. Many simulations and experimental investigations show that the above procedure is convergent, and leads to good spectral models. However, general mathematical proof of the convergence is not yet derived. The usefulness of the original algorithm has been proven among others in calibration laboratory [27, 28] to extend the bandwidth beyond technology limits, and in high voltage laboratory for non-destructive testing [3]. In the next section we will introduce a simplification, which results a significant improvement in computational speed, and we provide a convergence analysis.

V. Simplification of model based optimization of inverse filtering method

A. *Tikhonov's regularization*

If we choose Tikhonov's regularization method to calculate the inverse filter, the final deconvolution parameter(s) of our model based optimization method can be found without iteration as well, which speeds up the process.

Tikhonov proposes to regularize the ill-conditioned problem by introducing regularization operators beside the output error:

$$\|y - \hat{y}\| + \lambda\|\hat{x}\| \tag{10}$$

where $\|x\|$ is some kind of norm of x, e.g. Euclidean-norm, L1 norm, etc.

If we use (1), (2) and (3) formulae on (10), assume that the signal and the noise are uncorrelated, and minimize it using Euclidean norm, we get

$$H_{inv}(f) = \frac{\overline{H}(f)}{H(f)\overline{H}(f) + \lambda} = \frac{\overline{H}(f)}{|H(f)|^2 + \lambda} \tag{11}$$

Here the optimal value of λ is still unknown.

B. *Simplification*

Let us see, how the model based parameter optimization method can be simplified for the above special case of Tikhonov's regularization.

At the first step

$$|X_{m,0}(f)| = |Y(f)| \tag{12}$$

At the i-th step

$$|X_{m,i}(f)| = |Y(f)H_{inv_i}(f)| \tag{13}$$

Substitutuing the Tikhonov-solution into (13):

$$|X_{m,i}(f)| = |Y(f)\frac{\overline{H}(f)}{|H(f)|^2 + \lambda_i}| \tag{14}$$

where λ_i is the regularization parameter after the i-th iteration. By using this result in the $(i+1)$-th step, from (9), we can write the formula to be minimized (in λ_{i+1}) as

$$\sum_f |Y(f)|^2 \frac{|H(f)|^2}{(|H(f)|^2+\lambda_i)^2} \frac{\lambda_{i+1}^2}{(|H(f)|^2+\lambda_{i+1})^2} + \sum_f |N_m(f)|^2 \frac{|H(f)|^2}{(|H(f)|^2+\lambda_{i+1})^2} \tag{15}$$

The location of the minimum of this formula can be found, where the derivative by λ_{i+1} is zero:

$$\sum_f |Y(f)|^2 \frac{|H(f)|^2}{(|H(f)|^2+\lambda_i)^2} \frac{2|H(f)|^2\lambda_{i+1}}{(|H(f)|^2+\lambda_{i+1})^3} - \sum_f |N_m(f)|^2 \frac{2|H(f)|^2}{(|H(f)|^2+\lambda_{i+1})^3} = 0 \tag{16}$$

$$\lambda_{i+1} = \frac{\sum_f |N_m(f)|^2 \frac{|H(f)|^2}{(|H(f)|^2+\lambda_{i+1})^3}}{\sum_f |Y(f)|^2 \frac{|H(f)|^2}{(|H(f)|^2+\lambda_i)^2} \frac{|H(f)|^2}{(|H(f)|^2+\lambda_{i+1})^3}} \tag{17}$$

Assuming that at the end of the iteration $\lambda_{i+1} = \lambda_i$ in (17), we can simply write λ instead and the equation reduces to

$$\lambda = \frac{\sum_f |N_m(f)|^2 \frac{|H(f)|^2}{(|H(f)|^2+\lambda)^3}}{\sum_f |Y(f)|^2 \frac{|H(f)|^4}{(|H(f)|^2+\lambda)^5}} \tag{18}$$

$$\frac{\sum_f |N_m(f)|^2 \frac{|H(f)|^2}{(|H(f)|^2+\lambda)^3}}{\sum_f |Y(f)|^2 \frac{|H(f)|^4}{(|H(f)|^2+\lambda)^5}} - \lambda = 0 \tag{19}$$

The above two equations have the same solution, as the model based iterative method reported in the previous section, which requires an iteration to automatically extract spectral models of the signals, plus a numerical minimum finding in each iteration step. Solution of the proposed simplification requires only one numerical root finding (because λ cannot be expressed in closed form). However, it does not require additional iterations, the iterative model building process is reduced to one step. The root finding can be done e.g. with grid search, bisection method or successive approximation.

If we assume that the cost function is minimized on a grid in two steps (first on a rough grid, next on a fine grid around the minimum), both steps evaluated in 50-50 points, and the spectral models are refined in 20 consecutive steps, the cost function needs to be evaluated 100 x 20 = 2000 times with the original method. With the suggested modification it is reduced to about 100 (if we assume that the evaluation of the cost function requires nearly the same computation as evaluation of (18). The improvement is an order of magnitude. Our method can be further accelerated if the cost function is optimized by means of bisection (root finding of (19)) or successive approximation algorithms (iterative evaluation of (18)), which cannot be applied to the original model based algorithm. On that way we can reach an improvement of factor of 150.

C. Convergence analysis of numerical root finding methods

C.1 Grid search and bisection methods

For grid search and bisection method the end points of the search interval need to be determined. An obvious value for lower limit is 0, because λ - per definition - must always be ≥ 0. A reasonable upper limit for search can be $\max|H(f)|^2$. Above that value the distortion is extremely high, and the estimate is useless even if the the optimum is there. Grid search and bisection methods will be able to find an appropriate solution of (18) or (19) if the search interval contains the optimum, as these functions are continuous and smooth.

C.2 Successive approximation method

A further improvement in speed can be reached if we optimize (18) by means of successive approximation. However, this method has convergence limits, thus, there is a price for increased speed. It will be convergent if

$$\left|\frac{\partial}{\partial\lambda}\left(\frac{\sum_f |N_m(f)|^2 \frac{|H(f)|^2}{(|H(f)|^2+\lambda)^3}}{\sum_f |Y(f)|^2 \frac{|H(f)|^4}{(|H(f)|^2+\lambda)^5}}\right)\right| < 1 \tag{20}$$

This means that

$$\left|5\frac{\sum_f |N_m(f)|^2 \frac{|H(f)|^2}{(|H(f)|^2+\lambda)^3}}{\left(\sum_f |Y(f)|^2 \frac{|H(f)|^4}{(|H(f)|^2+\lambda)^5}\right)^2}\sum_f |Y(f)|^2 \frac{|H(f)|^4}{(|H(f)|^2+\lambda)^6} - 3\frac{\sum_f |N_m(f)|^2 \frac{|H(f)|^2}{(|H(f)|^2+\lambda)^4}}{\sum_f |Y(f)|^2 \frac{(|H(f)|^4}{(|H(f)|^2+\lambda)^5}}\right| < 1 \tag{21}$$

Let us examine the above equation at the borders (asymptotic values at $\lambda \to 0$ and $\lambda >> \max|H|^2$). For $\lambda >> \max|H|^2$, $(|H(f)|^2 + \lambda) \approx \lambda$. Applying this for (21), and simplifying it we get

$$\lambda < 0.5\frac{\sum_f |Y(f)|^2 |H(f)|^4}{\sum_f |N_m(f)|^2 |H(f)|^2} \tag{22}$$

This means that the starting value of λ should be smaller than very roughly the half of the signal-to-noise ratio, otherwise the successive approximation method could be divergent. Also, inherently lies in the equation that the signal-to-noise ratio should be "good enough" for convergence: if the noise energy is higher than the signal energy, the result of (22) is already < 0.5, but the corresponding optimal value of the regularization parameter is also high, thus the above condition is not fulfilled. (The solution for such a large SNR has only mathematical interest, it is usually useless for engineering purposes anyhow).

As $\lambda \to 0$, those frequencies dominate in the summations (let us collect these frequencies to f_0 set) , for which

$$\frac{(|H(f_0)|^2)^j}{(|H(f_0)|^2 + \lambda)^k} \tag{23}$$

is maximal. (Please note that the equation of convergence limit (21) consists similar fractions, but with different powers in the numerators and denominators. j and k in the above equation denote the power of the numerator and denominator.) The above equation reaches its maximum if

$$|H(f_0)|^2 = \frac{\lambda}{\frac{k}{j} - 1} \tag{24}$$

Substituting this in (21) we will get

$$5\frac{|N_m(f|_{H=0.71\lambda})|^2\frac{\lambda/2}{(3/2\lambda)^3}}{\left(|Y(f|_{H=1.22\lambda})|^2\frac{(2/3\lambda)^2}{(5/3\lambda)^5}\right)^2}|Y(f|_{H=0.71\lambda})|^2\frac{(\lambda/2)^2}{(3/2\lambda)^6}-3\frac{|N_m(f|_{H=0.58\lambda})|^2\frac{\lambda/3}{(4/3\lambda)^4}}{|Y(f|_{H=1.22\lambda})|^2\frac{(2/3\lambda)^2}{(5/3\lambda)^5}}<1 \tag{25}$$

$$13.612\frac{|N_m(f|_{H=0.71\lambda})|^2|Y(f|_{H=0.71\lambda})|^2}{|Y(f|_{H=1.22\lambda})|^2}-9.155\frac{|N_m(f|_{0.58\lambda})|^2}{|Y(f|_{H=1.22\lambda})|^2}<1 \tag{26}$$

The above derivation tells us the convergence limits of the successive approximation method. For reasonable starting values the successive approximation method is convergent; only at very small and very large λ values we might face divergent behavior. The suggested $max(|H|)$ value for starting point is a safe initial guess. Outside of the guaranteed convergence region bisection or grid search is a safe alternative to solve (18) or (19).

VI. Sensitivity analysis

Let us check the situation, if the noise level is incorrectly estimated, that is, all frequency components are over- or underestimated with the same amount:

$$\begin{aligned}|N_{est}(f)| &= (1+\epsilon)|N(f)| \\ |N_{est}(f)|^2 &= (1+2\epsilon+\epsilon^2)|N(f)|^2 \approx (1+2\epsilon)|N(f)|^2\end{aligned} \tag{27}$$

Equation (18) will be approximated. If the original equation is written in $\lambda = f(\lambda)$ form, the new equation will be

$$\begin{aligned}\lambda+\delta &\approx (1+2\epsilon)f(\lambda+\delta) \approx (1+2\epsilon)\left(\lambda+\frac{\mathrm{d}f(\lambda)}{\mathrm{d}\lambda}\delta\right) \\ \lambda+\delta &\approx \lambda+\frac{2\epsilon}{1-(1+2\epsilon)\frac{\mathrm{d}f(\lambda)}{\mathrm{d}\lambda}}\lambda \approx \left(1+\frac{2\epsilon}{1-\frac{\mathrm{d}f(\lambda)}{\mathrm{d}\lambda}}\right)\lambda\end{aligned} \tag{28}$$

(The multiplication by two appears, because N_{est} appears squared in the successive approximation form.)

(28) provides us a measure of sensitivity of noise model in our optimization algorithm. Luckily, it shows that λ estimate changes roughly with two times the error of the absolute value of noise model, which is tolerable.

VII. Verification

A. Verification on simulated data

The following simulation shows the correctness of the derived simplification of our proposed method.

We will also compare our optimization method with the ones that are the most promising techniques from the literature. The selection is based on the following criteria: the procedure needs to be fully automatic, no user interaction is required (thus [10, 15] is out of scope of comparison), the

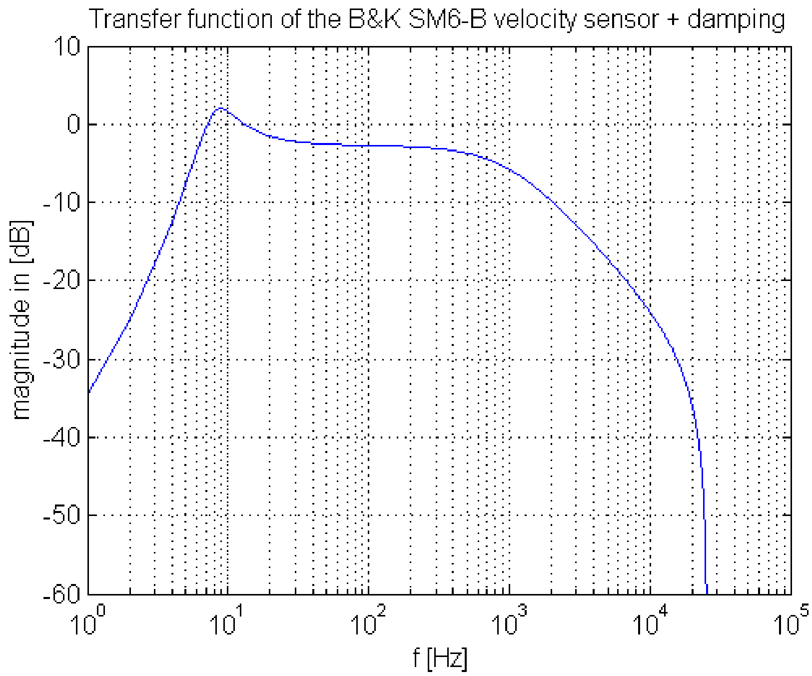

Fig. 2. Transfer function of the simulated system

method can be applied for Tikhonov type regularization (thus [6, 21] is out of scope of comparison), the method is not restricted to special signal types like e.g. step-like or Gaussian waveforms (thus [2, 19] is out of scope of comparison) and the method does not require much a priori knowledge about the signal to be measured (thus [20, 22–24] is out of scope of comparison). The convergence of the method in [4, 5] depends on implementation of FFT; we couldn't reproduce their results. There are two methods remaining, which can be compared to ours on an objective manner, namely Morozov's technique [11] and the classical L-curve method [16, 17].

As distortion the characteristic of a Bruel&Kjær velocity sensor is used, combined with a first order low-pass filter, having a corner frequency of 1 kHz, as additional damping. The IIR filter coefficients, assuming 51.2 kHz sampling frequency[1], are:

$$\begin{aligned} a_{sensor} &= [1.0 \text{ -}1.99939 \ 0.999396]; \\ b_{sensor} &= [0.730889 \text{ -}1.46166 \ 0.730777]; \\ a_{damping} &= [1.0 \text{ -}0.884239]; \\ b_{damping} &= [0.0578803 \ 0.0578803]; \end{aligned} \tag{29}$$

The transfer characteristic can be seen in Fig. 2. $\max(H(f))$ is 1.2539. The input- and the distorted, noisy output signals are depicted in Fig. 3.

Signal models of model based algorithm are depicted in Fig. 4. Noise level for the noise model is extracted from the spectrum of the output measurement by averaging the square of its absolute value at high frequency band. Absolute value of the signal model is derived from the regularized estimate of the input signal.

[1] This is the sampling frequency of a vibration monitoring system developed by Dr. Bakó et al.

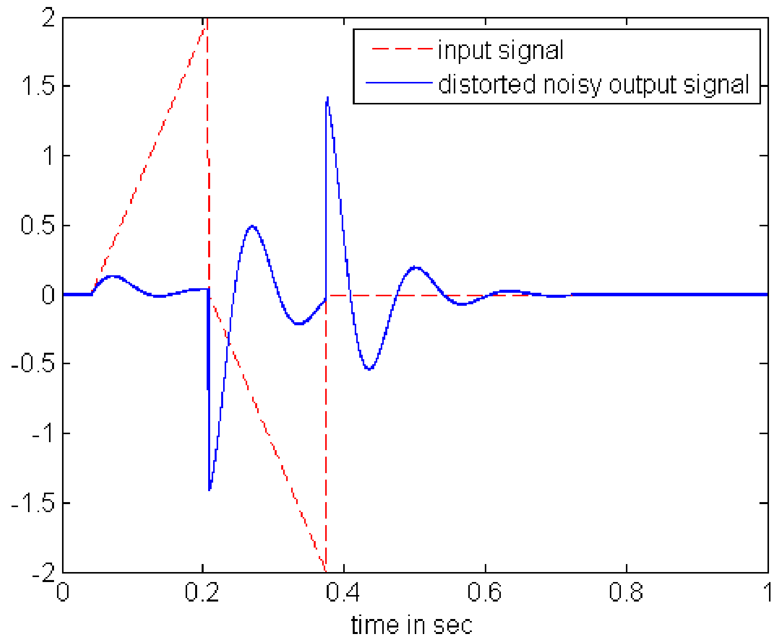

Fig. 3. Input and output signals of the simulation

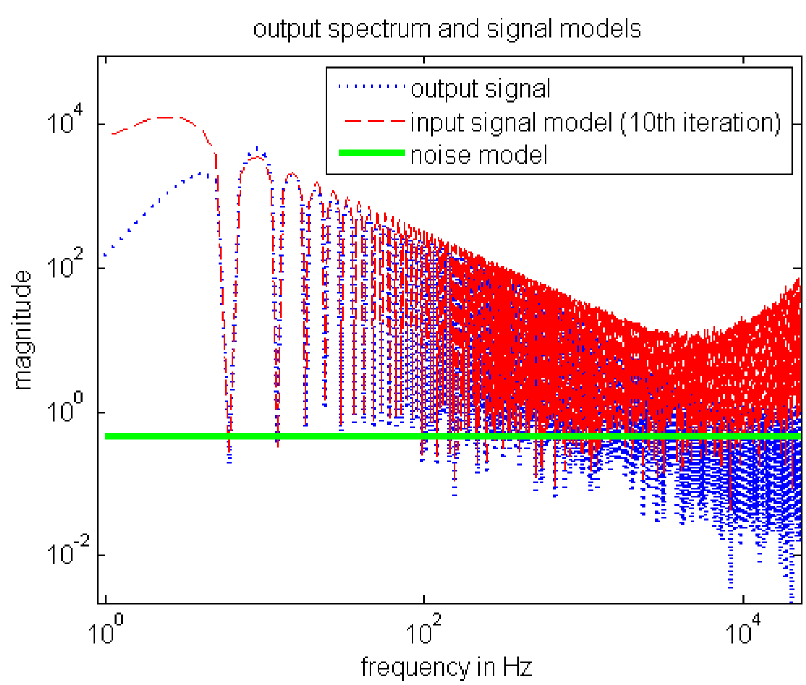

Fig. 4. Spectrum of the simulated output signal of the system together with the spectral models of the signal and noise

A.1 Performance comparison

As in the simulation the true input signal is known, the exact optimal λ parameter, "true optimum", can be determined. In the following table the true optimum and the estimate values for different methods can be found. The RMS errors between the original input signal and the restored one (input error) are also calculated. The simulations were repeated on the same noise level, but with different noise records to check stability (last column of table).

	λ	RMS error	rel. std. of RMS error
True optimum:	6.9847E-5	16.45	1
Model-based, original m.	7.4702E-5	16.47	1.0269
Model-based, new bisec m.	7.4702E-5	16.47	1.0269
Model-based, new succ. appr.	7.4702E-5	16.47	1.0269
Morozov's method	1.7655E-4	20.49	1.3400
L-curve method	1.5399E-7	74.51	263.79

It can be seen that the three estimation methods of model based approach provide the same regularization parameter, which is close to the optimum. The differences between the RMS errors are insignificant, thus the speed improvement did not affect performance. Solutions are close to the exact optimum with both the original model based method and the improved speed algorithm. Morozov's method performed also well. L-curve optimization, however, for this type of signal and system (bandpasss nature) systematically under-regularized the solution. The reason might be that the optimality criterion is heuristic.

If we use the fast successive approximation method to solve (18) care should be taken with respect to convergence limits. To examine its convergence properties, the derivative of (18) is calculated for a large range of λ, which is depicted in Fig. 5. The absolute value of this derivative should be smaller than 1 for convergence (in the case of succ. approx.). The upper convergence limit according to (22) is 4.1939E+05. The lower limit is at 2.5119e-05. Outside of this convergence region there is still a chance for convergence, but it is not guarantied.

A.2 Noise sensitivity

The next step is to check the performance on different noise levels. Fig. 6 shows input error relative to the optimal one. Both the model-based and Morozov's algorithm perform well. At the SNR range of 30..70 dB (which is the most common in engineering practice in the case of inverse filtering of deterministic signals) the model based optimization outperforms Morozov's. The L-curve method is unstable, only in the region of 5..15 dB showed comparable solutions. The estimate is acceptable in the region of 3..40 dB, however, the input RMS error is 5 times that of the optimal solution. Outside of this region the estimate tends to amplify the noise on an unacceptable level.

Let us check the sensitivity of the model based method for the estimate of noise level. The derivative of (18) at the λ selected by our algorithm is -0.39 in the simulation. If the noise estimate is multiplied by 1.1 or 0.9, then, according to the perturbation calculations (see (28), the selected λ would be multiplied by $1 \pm 0.1 * \frac{2}{1.3971} = 1 \pm 0.1432$. The simulation provided the following optimal λ values, which correspond to the expectations of the approximations in (28):

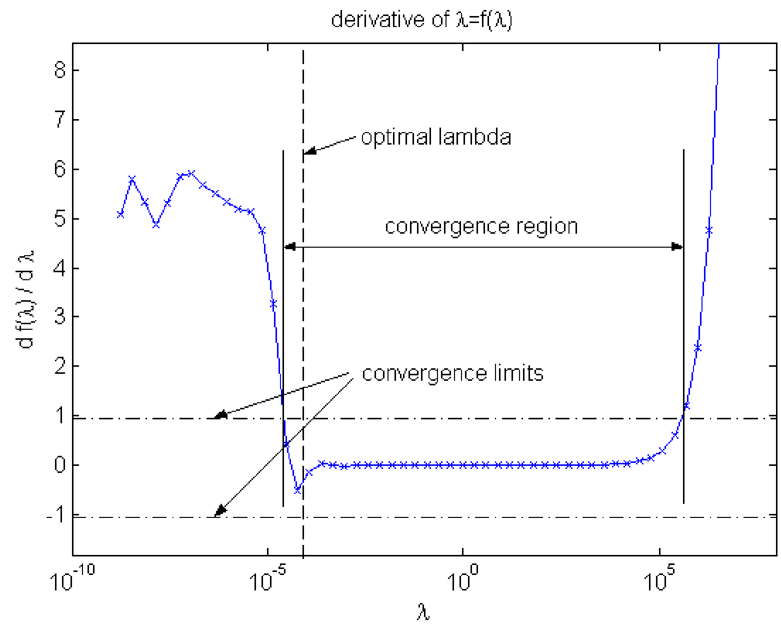

Fig. 5. Convergence limits of successive approximation method. Derivative of (18) for several λ values, which should be between -1 and +1 for guarantied convergence.

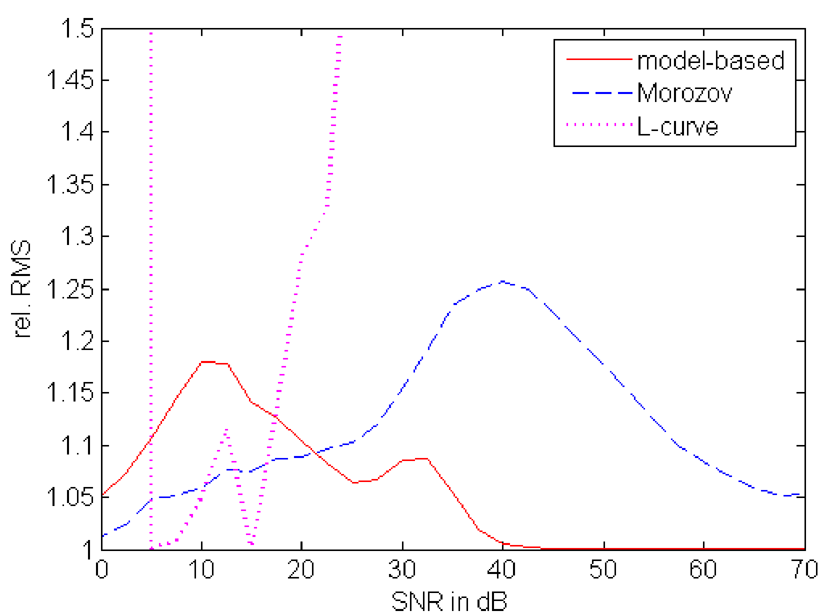

Fig. 6. Input RMS error relative to the RMS error of the optimal solution as a function of Signal-to-noise ratio (SNR)

N_{est} multiplicator	λ	Relative difference
1.0	7.4702E-005	0.0
1.1	8.5670E-005	0.1468
0.9	6.4224E-005	-0.1403

Let us check the effect of noise level estimate on the input error, since this is the primary measure of the performance. We compared the model based method with Morozov's algorithm (L-curve method does not use noise level estimate).

	Input error (RMS) relative to optimal one	
N_{est} multiplicator	model-based	Morozov
0.9	1.0017	1.1717
1.0	1.0011	1.2454
1.1	1.0109	1.3225

The model-based method seems to be much more robust with respect to error of noise level estimate. (Actually, the primary drawback of Morozov's method is its strong dependence on this a priori information.)

A.3 Computation speed

The computation speed of the model based approach has been significantly improved. We used MATLAB as simulation framework on a PC having a 1.7 GHz Intel Core i5 quad-core CPU, and measured a roughly 70 times improvement for bisection method, and a 150 times improvement for successive approximation. For a relatively large record length of 50000 samples optimization with successive approximation took around 0.5 sec, compared to 46 sec (original model based algorithm). Morozov's algorithm with its simplicity is very fast (0.06 sec), however, not as robust, as it was shown. L-curve method requires more computation (3 sec). Usually we are going to achieve a computation time below 1 sec to provide a convenient environment for the user. With the proposed speed improvement the model based optimization algorithm is within this range even for a large dataset, with the additional benefit of its robustness and systematic derivation from the input error criterion.

B. Verification on measured data

Let us check the usefulness of the proposed method on measurement data. We analyzed the finite bandwidth of an accelerometer (ADXL203 as Device Under Test (DUT)). This is a low bandwidth MEMS sensor, based on the principle of unbalanced differential capacitors due to acceleration. The sensor has been excited by a shaker. As a reference measurement, a Bruel&Kjær high bandwidth precision accelerometer has been used (type 4399, miniature piezoelectric accelerometer). The transfer function of DUT has been modeled by a first order system. Model parameters have been fitted to measurements at many frequencies for sinusoidal excitation. The test system has been excited with a short pulse to the shaker, having a pulse width of approximately 1.5 ms. The shaker responds to this excitation with a damped oscillation. Fig. 7 shows the output of the accelerometers. Deconvolution without any regularization yields a large noise amplification with useless result (Fig. 8). Optimal reconstruction with the proposed algorithm is shown in Fig. 9, where the reconstructed curve perfectly fits the reference measurement.

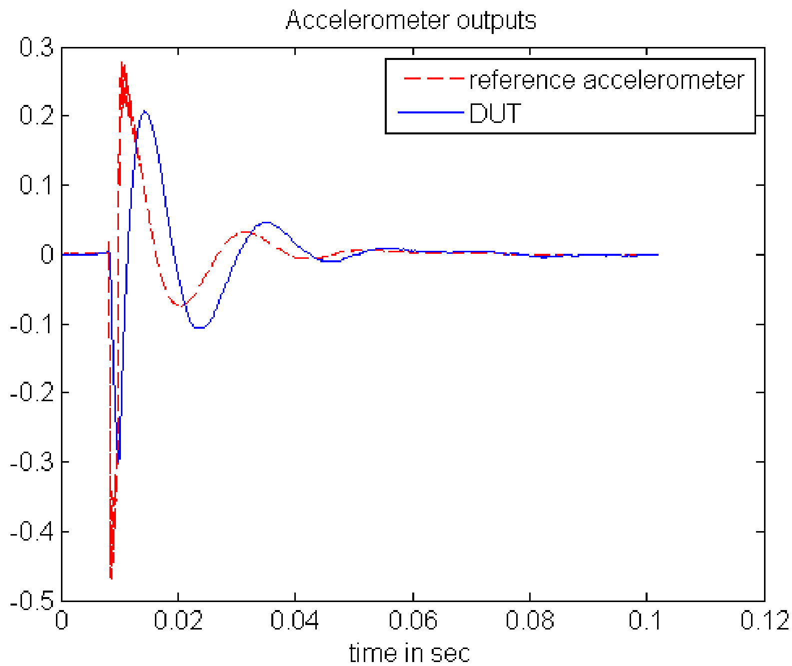

Fig. 7. Measured accelerometer signals. Solid line: DUT (ADXL203), dashed line: reference accelerometer

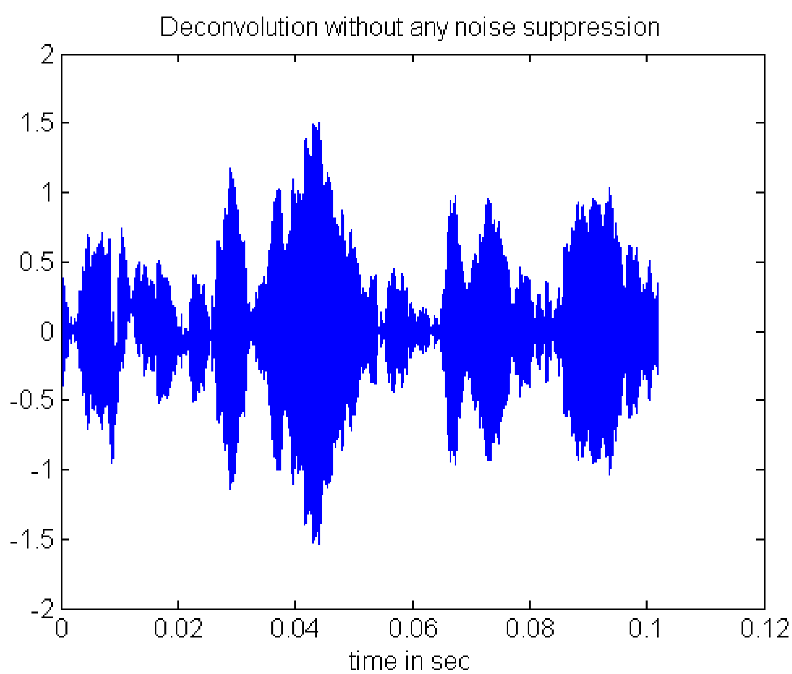

Fig. 8. Reconstruction of the accelerometer signal without regularization

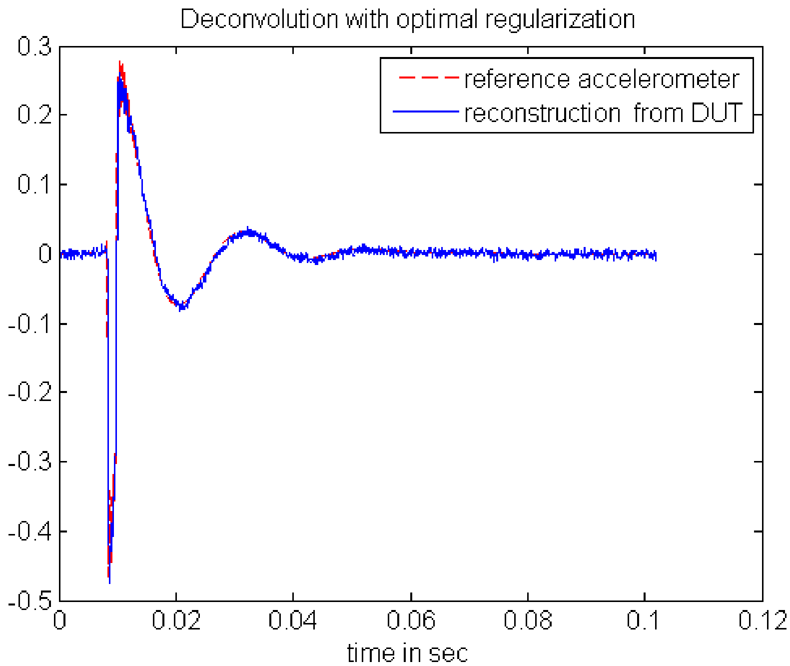

Fig. 9. Reconstruction of the accelerometer signal with the proposed algorithm. Solid line: reconstruction, dashed line: reference accelerometer. The two signals are nearly the same.

VIII. Conclusion

In this paper a model based optimization technique for ill-posed inverse filtering problems, developed by one of the authors, was analyzed. This technique optimizes parameters of parametric deconvolution methods. Its advantage is that the error function to be minimized is directly derived from the input error criterion with some approximations and spectral modeling. Spectral models are built automatically. As a novelty, it was shown that the original method, which contains a minimum search and an iteration, can be simplified to a root finding problem, if Tikhonov type regularization is used, and thus the method can be significantly facilitated (speed improvement of two order of magnitudes). The root finding can be calculated with bisection- or with successive approximation methods as well. The convergence properties of the successive approximation method was also analyzed. Robustness of the model-based method has been checked with respect to noise level and error of noise level estimate. Both simulation and measurement data showed that the estimates are correct. A further improvement could be the extension of this novel method to multiparameter optimization, and to other types of regularization methods.

Acknowledgments

This work was partially supported by the ARTEMIS JU and the Hungarian National Research, Development and Innovation Fund in the frame of the R5-COP project, and Hungarian Scientific Research Fund (OTKA, 115820).

References

[1] A. N. Tikhonov and V. Y. Arsenin, *Solution of ill-posed problems*, John Wiley & Sons, Inc., New York, 1977.

[2] M . Bertocco, C. Narduzzi, C. Offelli and D. Petri, *An improved method for iterative identification of bandlimited linear systems*, IEEE Instrumentation and Measurement Technology Conference, Atlanta (GA), May 14-16, 1991, CH2940-5/91, pp. 368-372.

[3] T. Dabóczi and I. Kollár, *Multiparameter optimization of Inverse Filtering Algorithms*, IEEE Trans. on Instrumentation and Measurement, Vol. 45, No. 2, 1996, pp. 417-421.

[4] N. S. Nahman and M. E. Guillaume, *Deconvolution of Time Domain Waveforms in the Presence of Noise*, National Bureau of Standards, Tech. Note 1047, NBS, Boulder, CO. USA.

[5] N. S. Nahman, *Software correction of measured pulse data*, in Fast Electrical and Optical Measurements, Ed. by T.E Thompson and L. H. Luessen, NATO ASI Series, Martin Nijhoff Publishers, Dordrecht, The Netherlands, 1986, pp. 351-417.

[6] N. H. Younan, A. B. Kopp, D. B. Miller and C. D. Taylor, *On correcting HV impulse measurements by means of adaptive filtering and deconvolution*, IEEE Trans. on Power Delivery, Vol. 6, No. 2, 1990, pp. 501-506.

[7] P. B. Crilly, *A The use of a cross-correlation technique to enhance Jansson's deconvolution procedure*, Journal of Chemometrics, Vol. 4, No. 4, July 1990, pp. 291298.

[8] P. B. Crilly, *A quantitative evaluation of various iterative deconvolution algorithms*, IEEE Trans. on Instrumentation and Measurement, Vol. 40, No. 3, 1991, pp. 558-562.

[9] Paul Benjamin Crilly, Andre Bernardi, Peter A. Jansson and Luiz Eduardo Borges da Silva, *Improving the Convergence Rate of Jansson's Deconvolution Method*, IEEE Trans. on Instrumentation and Measurement, Vol. 51, No. 6, Dec. 2002, 1142-1144.

[10] B. Parruck and S. M. Riad, *An optimization criterion for iterative deconvolution*, IEEE Trans. on Instrumentation and Measurement, Vol. 32, No. 10, 1983, pp. 137-140.

[11] Morozov, V. A., *Methods for Solving Incorrectly Posed Problems* Springer, New York, 1984.

[12] S. K. Lehman, *Deconvolution Using a Neural Network*, NTIS Report, NTIS No. DE91007114/HDM, Report No. UCRL-ID-195439.

[13] Philippe Neveux, E. Sekko, and Gerard Thomas, *A Constrained Iterative Deconvolution Technique with an Optimal Filtering: Application to a Hydrocarbon Concentration Sensor*, IEEE Trans. on Instrumentation and Measurement, Vol. 49, No. 4, Aug. 2000, 852-856.

[14] Claudio Narduzzi, *Inverse Filtering With Signal-Adaptive Constraints*, IEEE Trans. on Instrumentation and Measurement, Vol. 54, No. 4, Aug. 2005,pp. 1553-1559.

[15] T. Dhaene, L. Martens and D. De Zuttert, *Generalized iterative frequency domain deconvolution technique*, IEEE Instrumentation and Measurement Technology Conference, Irvin, Orange Country, California, May 18-20, 1993, 93CH3291-0, pp. 85-87.

[16] Per Christian Hansen, *Analysis of Discrete Ill-Posed Problems by Means of the L-Curve*, SIAM Review, Vol. 34, No. 4, Dec. 1992, pp. 561-580.

[17] P. C. Hansen and D. P. OLeary, *The use of the L-curve in the regularization of discrete ill-posed problems*, SIAM J. Sci. Comput., Vol. 14, No. 6, 1993., pp. 14871503.

[18] Chang Liu, Lijun Xu, Zhang Cao and Hugh McCann, *Reconstruction of Axisymmetric Temperature and Gas Concentration Distributions by Combining Fan-Beam TDLAS With Onion-Peeling Deconvolution*, IEEE Trans. on Instrumentation and Measurement, Vol. 63, No. 12, Dec. 2014, pp. 3067-3075.

[19] Sandip Roy and Michael Souders, *Noniterative Waveform Deconvolution Using Analytic Reconstruction Filters With Time-Domain Weighting*, IEEE Trans. on Instrumentation and Measurement, Vol. 50, No. 4, Aug. 2001, pp. 899-904.

[20] T. Olofsson, *Deconvolution and Model-Based Restoration of Clipped Ultrasonic Signals*, IEEE Trans. on Instrumentation and Measurement, Vol. 54, No. 3., 2005, pp. 1235-1240.

[21] D. Szolgay and T. Sziranyi, *Optimal stopping condition for iterative image deconvolution by new orthogonality criterion*, Electronics Letter, doi: 10.1049/el.2010.3692, Vol. 47, No. 7., 31st March 2011.

[22] M. Kido, Y. hirata, S. Yamada and K. Kondo, *Automatic Restoration of X-ray Photography by Estimating Point Spread Function*, Proceedings of the IEEE International Symposium on Communications and Information Technologies (ISCIT-2010), 26-29 Oct. 2010, Tokyo, Japan, pp. 1180-1184.

[23] A. Loktyushin and S. Harmeling, *Automatic foreground-background refocusing*, IEEE International Conference on Image Processing, 11-14 Sept. 2011, Brussels, Belgium, 2011, pp. 3445-3448.

[24] L. Thibon, F. Soulez and E. Thiebaut, *Fast Automatic Myopic Deconvolution of Angiogram Sequences*, IEEE 11th International Symposium on Biomedical Imaging (ISBI), April 29 - May 2 2014, Beijing, China, 2014, pp. 1067-1070.

[25] P. B. Crilly, *Error Analysis with Deconvolution Algorithms*, IEEE Trans. on Instrumentation and Measurement, Vol. 42, No. 1, 1993, p. 78.

[26] A. Rosenfeld and A. C. Kak, *Digital Picture Processing*, Second Edition, Volume 1, Academic Press, Inc., Orlando, Florida, 1982.

[27] Deyst J. P., N. G. Paulter, Dabóczi T., Stenbacken G. N. and Souders T. M., *A Fast Pulse Oscilloscope Calibration System*, IEEE Trans. on Instrumentation and Measurement Vol. 47, No. 5, 1998, pp. 1037-1041.

[28] Dabóczi T. and Bakó T. B., *Inverse Filtering of Optical Images*, IEEE Trans. on Instrumentation and Measurement, Vol. 50, No. 4, 2001, pp. 991-994.

Tamás B. Bakó was born in Budapest, Hungary, in 1976. He received M.Sc. degree in electrical engineering in 1999 and a second degree in biomedical engineering in 2001 from the Budapest University of Technology and Economics. He received the Ph.D degree in 2005 at the same department. He spent 9 years in the industry on turbine monitoring projects (accurate measurements, noise reduction and inverse filtering) at Evosoft Hungary Kft. Presently, he works as quantitative developer at Morgan Stanley.

Tamás Dabóczi (M'98, SM'04) received the M.Sc. and Ph.D. degrees in electrical engineering from Budapest University of Technology, Budapest, Hungary, in 1990 and 1994, respectively. In 1990 he joined the same university. He spent several months at the Swiss Federal Institute of Technology (ETH), Zrich, Switzerland as academic guest in 1992; at the Technical University of Karlsruhe, Germany as visiting researcher in 1994. In 1995 he was invited to National Institute of Standards and Technology, NIST, USA, Gaithersburg, MD as guest researcher, where he developed and implemented inverse filtering algorithm for calibration of ultra high-speed digital sampling oscilloscopes. Currently he is the head of Department of Measurement and Information Systems, Budapest University of Technology and Economics, Budapest, Hungary. His research area includes embedded systems, cyber-physical systems, digital signal processing, especially inverse filtering. He is senior member of IEEE.