# COMPUTATIONALLY EFFICIENT NORM-CONSTRAINED ADAPTIVE BLIND DECONVOLUTION USING THIRD-ORDER MOMENTS

Patrik Pääjärvi and James P. LeBlanc

Luleå University of Technology SE-971 87 Luleå SWEDEN

{patrik,leblanc}@csee.ltu.se

# ABSTRACT

Third-order central moments have been shown to be well suited as objective functions for blind deconvolution of impulsive signals. Online implementations of such algorithms may suffer from increasing filter norm, forcing adaptation under constrained filter norm. This paper extends a previously known efficient algorithm with selfstabilizing properties to the case of using a third-order moment objective function. New results herein use averaging analysis to determine adaptation stepsize conditions for asymptotic stability of the filter norm.

### 1. INTRODUCTION



Fig. 1. Model of general adaptive blind deconvolution.

Blind deconvolution is used for identification or equalization of unknown systems in situations where only the system output can be observed. A general discrete-time model is shown in Figure 1, where n denotes a time index,  $s_n$  the unknown source, and **c** the unknown system. The object is to find the deconvolution filter **f** that approximately inverts the unknown system, so that  $y_n$  becomes an estimate of  $s_n$ .

If the deconvolution filter  $\mathbf{f}$  is iteratively adjusted according to some *error function*, we get an *adaptive* blind deconvolution setting. The error function (corresponding to the error signal of the standard LMS algorithm) is related to the gradient of an *objective function* of the filter output  $y_n$ . Adaptation of  $\mathbf{f}$  is aimed at maximizing the objective function.

Minimum Entropy Deconvolution (MED) methods [1], [2] are based on the idea that; given an uncorrelated sequence  $s_n$ , the probability distribution of  $u_n$  is closer to a Gaussian distribution compared to that of  $s_n$ . This consequence of the central limit theorem allows for blind deconvolution based on discriminating the distribution of  $y_n$  from a Gaussian distribution. An objective function for adaptive MED should therefore be a measure of 'how Gaussian'  $y_n$  is. *Higher-order moments* (order greater than two) are popular measures of Gaussianity, especially the *kurtosis* (normalized fourthorder moment).

Apart from their ability to measure Gaussianity, higher-order moments can also be used to describe how heavy-tailed the probability density function (PDF) of a signal is. A signal with a heavy-tailed PDF has a 'spiky' appearance. This type of distribution characterization allows for blind deconvolution without the assumption of the source signal necessarily being a white sequence.

If  $s_n$  is known to have a non-zero third-order moment, this *asymmetry* allows for exploitation of *skewness* as an objective function, as an alternative to kurtosis. The skewness of a stochastic variable x is the normalized third-order moment

$$S(x) = \frac{\mathrm{E}\{x^3\}}{(\mathrm{E}\{x^2\})^{3/2}},\tag{1}$$

where  $E\{\cdot\}$  denotes expectation. Since all odd-order moments of a signal with symmetric PDF are zero, the use of odd-order moments such as (1) is restricted to asymmetric signals.

In previous work, skewness has been used for blind deconvolution of impulsive signals (i.e. asymmetric signals dominated by positive 'spikes'). When compared to kurtosis, skewness generally gives faster convergence of algorithms, and is less sensitive to additive white Gaussian noise [3], [4]. This motivates why exploitation of signal asymmetry using skewness may be preferable to kurtosisbased methods.

Due to the relative complexity of its gradient equation, (1) may not be suitable as an objective function for real-time applications requiring minimal computational cost. A more computationally efficient function is

$$\mathcal{O}(x) = \frac{1}{3} \mathrm{E}\left\{x^{3}\right\},\tag{2}$$

a scaled version of the third-order moment of the stochastic variable x. While easier to estimate than skewness, (2) is not scale invariant in x. That is,  $\mathcal{O}(x) \neq \mathcal{O}(kx)$  for  $k \neq 1$ . As a consequence, standard gradient-ascent algorithms based on (2) give a rapid increase in filter norm over iterations. In fact, such problems arise for general choices of objective functions when impulsive signals are deconvolved [5]. Increasing filter norm causes numerical problems in

This work was partially supported by Vinnova's Vinnväxt program ProcessIT Innovations.

implementations, especially on fixed-point architectures. Therefore, a blind deconvolution algorithm maximizing the third-order moment must work under *constrained filter norm*.

An overview of several norm-constrained gradient adaptation algorithms can be found in [6] and [7], mainly considering objective functions of the form

$$J = \pm \frac{1}{p} \mathbb{E}\left\{ |y_n|^p \right\},\,$$

where p is a positive integer. Since these functions are sign-invariant in their argument, they are unable to exploit asymmetry.

In this paper, one of the algorithms from [6] and [7] is studied when (2) is the specific function to be maximized. The work in [6] and [7] is extended, using averaging analysis, to determine conditions for asymptotic numerical stability. The computational cost of this algorithm is compared to other commonly employed methods.

# 2. NOTATION

Referring to Figure 1, the deconvolution filter  $\mathbf{f}$  is an adaptive, real FIR filter of order N, represented at time n by its coefficient vector  $\mathbf{f}_n \triangleq [f_{0n} f_{1n} \cdots f_{Nn}]^{\mathrm{T}}$ . The norm  $\|\mathbf{f}_n\|$  of  $\mathbf{f}_n$  is defined as the Euclidean, or  $\ell_2$ -norm. Denoting the filter regressor by the vector of real samples  $\mathbf{u}_n \triangleq [u_n \ u_{n-1} \ \cdots \ u_{n-N}]^{\mathrm{T}}$ , the filter output becomes the vector inner product  $y_n = \mathbf{f}_n^{\mathrm{T}} \mathbf{u}_n$ .

The objective function to be maximized is the third-order moment of the filter output  $y_n$ ,

$$\mathcal{O}(y_n) \triangleq \frac{1}{3} \mathbb{E}\left\{y_n^3\right\} = \frac{1}{3} \mathbb{E}\left\{\left(\mathbf{f}_n^{\mathrm{T}} \mathbf{u}_n\right)^3\right\}.$$
(3)

Throughout the remainder of this paper, the operation count associated with implementations of each of the presented algorithms are taken under the assumption that all expectations of the form  $E\{x_n\}$  are estimated by instantaneous values  $x_n$ , as is customary for on-line applications.

# 3. ADAPTATION UNDER CONSTRAINED FILTER NORM

### 3.1. Adaptation Using Steepest Ascent

Adaptation by steepest ascent is used to adjust the filter to maximize the objective (3),

$$\mathbf{f}_{n+1} = \mathbf{f}_n + \mu \, \boldsymbol{\nabla}_n,\tag{4}$$

where  $\mu$  is a small positive stepsize and  $\nabla_n$  is the gradient of  $\mathcal{O}$  with respect to  $\mathbf{f}_n$ ,

$$\boldsymbol{\nabla}_{n} \triangleq \frac{\partial \mathcal{O}}{\partial \mathbf{f}_{n}} = \frac{\partial \mathcal{O}}{\partial y_{n}} \frac{\partial y_{n}}{\partial \mathbf{f}_{n}} = \mathrm{E}\left\{y_{n}^{2} \mathbf{u}_{n}\right\}.$$
(5)

Using (4),  $\mathbf{f}_n$  is iteratively adjusted until  $\mathcal{O}$  attains a maximum. Note that for any number  $\alpha$  and any vector  $\mathbf{f}$ ,

$$\mathcal{O}(\alpha \mathbf{f}) = \alpha^3 \mathcal{O}(\mathbf{f}).$$

Hence, for any filter vector  $\mathbf{f}$ , we can improve  $\mathcal{O}$  with the vector  $\alpha \mathbf{f}$ ,  $\alpha > 1$ . Therefore (4) will never converge since  $\nabla_n$  will never approach zero. Instead, the norm of  $\mathbf{f}_n$  will rapidly increase over

iterations. A simple way to deal with this is to combine (4) with a frequent normalization procedure,

$$\mathbf{f}_{n+1} \leftarrow \frac{\mathbf{f}_{n+1}}{\|\mathbf{f}_{n+1}\|}.$$
(6)

While this would keep  $\|\mathbf{f}_n\| = 1$  over iterations, the computational cost associated with combining (4) and (6) is relatively large; on the order of 4N operations per iteration for an  $N^{\text{th}}$ -order filter. Therefore, alternative ways to do steepest ascent under constrained filter norm are desired.

#### 3.2. Orthogonal Gradient Decomposition

Recognize that a scaling  $\alpha \mathbf{f}_n$  only results in a scaling  $\alpha y_n$  of the filter output signal; the 'quality' of deconvolution is not affected. A reasonable approach would therefore be to avoid updating  $\mathbf{f}_n$  in the radial direction.

Consider a decomposition of  $\nabla_n$  into  $\nabla_n = \mathbf{R}_n + \mathbf{P}_n$ , where  $\mathbf{R}_n$  is the orthogonal projection of  $\nabla_n$  onto  $\mathbf{f}_n$ ,

$$\mathbf{R}_{n} \triangleq \frac{\boldsymbol{\nabla}_{n}^{\mathrm{T}} \mathbf{f}_{n}}{\|\mathbf{f}_{n}\|^{2}} \mathbf{f}_{n}.$$
(7)

Then modify the steepest-ascent algorithm to only update  $\mathbf{f}_n$  in non-radial directions,

$$\mathbf{f}_{n+1} = \mathbf{f}_n + \mu \, \mathbf{P}_n = \mathbf{f}_n + \mu \left[ \boldsymbol{\nabla}_n - \mathbf{R}_n \right]. \tag{8}$$

This algorithm can be viewed as a search for local maximum points of the objective function in the *tangent space of the hypersphere*  $\|\mathbf{f}\| = \|\mathbf{f}_n\|$  at  $\mathbf{f} = \mathbf{f}_n$ . Unlike the standard algorithm (4), the modified version is expected to converge to points at which  $\mathbf{P}_n$  approaches zero.

Ideally, the search for local maximum points should be restricted to some hypersphere,  $\|\mathbf{f}_n\| = C$ , to ensure that the filter norm stays fixed. For (8), it is straightforward to show that  $\|\mathbf{f}_{n+1}\| \ge \|\mathbf{f}_n\|$ . Hence, although the growth in  $\|\mathbf{f}_n\|$  will not be as rapid as for the standard algorithm, this modified gradient ascent must be combined with an infrequent normalization of  $\mathbf{f}_n$ . Even without normalization, the operation count per iteration for an implementation of (8) is on the order of 4N for an  $N^{\text{th}}$ -order filter. Hence, this algorithm offers no computational savings.

#### 3.3. Pseudo-Orthogonal Gradient Decomposition

A slight modification of (8) is achieved if the factor  $1/||\mathbf{f}_n||^2$  is neglected in (7). Define

$$\widetilde{\mathbf{R}}_n \triangleq \left( \mathbf{\nabla}_n^{\mathrm{T}} \mathbf{f}_n \right) \mathbf{f}_n \quad \text{and} \quad \widetilde{\mathbf{P}}_n \triangleq \mathbf{\nabla}_n - \widetilde{\mathbf{R}}_n,$$

and do the filter adaptation as

$$\mathbf{f}_{n+1} = \mathbf{f}_n + \mu \widetilde{\mathbf{P}}_n$$
$$= \left(1 - \mu \nabla_n^{\mathrm{T}} \mathbf{f}_n\right) \mathbf{f}_n + \mu \nabla_n.$$
(9)

As noted in [6] and [7], if  $\nabla_n^T \mathbf{f}_n > 0$ , this algorithm operates in a stable manner, maintaining approximately unit filter norm. Absolute convergence of the algorithm will, however, ultimately depend on  $\mu$ .

From the definitions of the gradient (5) and the objective function (3), we find that

$$\boldsymbol{\nabla}_{n}^{\mathrm{T}}\mathbf{f}_{n} = \mathrm{E}\left\{y_{n}^{3}\right\} = 3\,\mathcal{O}(y_{n}),\tag{10}$$

i.e. the quantity  $\nabla_n^{\mathrm{T}} \mathbf{f}_n$  is proportional to the objective function to be maximized by the algorithm. Although  $\mathcal{O}(y_n) > 0$  cannot be guaranteed for all *n*, the algorithm will most likely tend towards a positive objective over iterations for a well-conditioned problem.

Using (10), (9) can be rewritten as

$$\mathbf{f}_{n+1} = \left(1 - \mu \operatorname{E}\left\{y_n^3\right\}\right) \mathbf{f}_n + \mu \, \boldsymbol{\nabla}_n,\tag{11}$$

which exposes the algorithms simplicity. The computational cost of implementing this algorithm with an  $N^{\text{th}}$ -order filter is on the order of 3N operations per iteration. Furthermore, (11) contains only multiplications and additions (i.e. no divisions), making it highly suitable for implementation on fixed-point digital signal processors, which are specialized at performing such arithmetic operations.

The next section considers the asymptotical behavior of this algorithm and derives a sufficient condition on  $\mu$  for numerical stability.

# 4. ASYMPTOTIC STABILITY OF THE PSEUDO-ORTHOGONAL GRADIENT DECOMPOSITION ALGORITHM

To analyze the behavior of  $\|\mathbf{f}_n\|$  over iterations in the algorithm (11), define

$$\varepsilon_n \triangleq \|\mathbf{f}_n\|^2 - 1$$
 (12)

as the deviation of  $||\mathbf{f}_n||^2$  from unity at time *n*. Multiplying both sides of (11) with their transposes and subtracting off one, gives after rearranging terms

$$\varepsilon_{n+1} = \left(1 - \mu 2 \mathbb{E}\left\{y_n^3\right\}\right) \varepsilon_n + \mu^2 \|\widetilde{\mathbf{P}}_n\|^2.$$
(13)

This expression describes how the norm of  $\mathbf{f}_n$  deviates from unity over iterations. The goal is to derive sufficient conditions on  $\mu$  such that  $\varepsilon_n \to 0$  as  $n \to \infty$ .

Note that (13) is a difference equation of the form

$$\varepsilon_{n+1} = \varepsilon_n + \mu g(n, \varepsilon_n, \mu),$$
 (14)

where g is a nonlinear, stochastic and time-varying function. Assuming that the stepsize  $\mu$  is small, (14) may be approximated by the *averaged system* 

$$\overline{\varepsilon}_{n+1} = \overline{\varepsilon}_n + \mu g_{av}(\overline{\varepsilon}_n), \tag{15}$$

where

$$g_{av}(\varepsilon) = \mathbf{E} \left\{ g(n,\varepsilon,0) \right\} \Big|_{\varepsilon = \text{constant}} .$$
(16)

The necessary conditions for the approximation of (14) with the averaged system (15) to be valid are essentially that, over a fixed time interval;  $g_{av}$  is time invariant,  $|\overline{\varepsilon}|$ ,  $|g_{av}|$  are bounded, and g and the difference  $g - g_{av}$  fulfill global Lipschitz conditions in  $\varepsilon$  and  $\mu$ . Refer to [8, Ch. 9] for details.

Although the expectation in (16) is taken with  $\mu = 0$ , we choose to regard  $g_{av}$  as a function of both  $\varepsilon$  and  $\mu$  to investigate how the stepsize affects the asymptotical behavior of the algorithm. Comparing (13) with (14) gives

$$g(n,\varepsilon_n,\mu) = -2\mathrm{E}\left\{y_n^3\right\}\varepsilon_n + \mu \|\tilde{\mathbf{P}}_n\|^2$$

and the averaged system from (16) as

$$g_{av}(\varepsilon,\mu) = -2\,S_y\,\varepsilon + \mu\,\widetilde{P}^2,$$

where

$$S_y \triangleq \mathrm{E}\left\{y_n^3\right\},\tag{17}$$

$$\widetilde{P}^2 \triangleq \mathrm{E}\left\{\|\widetilde{\mathbf{P}}_n\|^2\right\}$$
(18)

are assumed to be time invariant.

While the assumptions of time-invariance in (17) and (18) are not realistic over a larger span of iterations (in fact, note that  $S_y$  is proportional to the objective function to be maximized), (17) and (18) are approximately time invariant over limited number of iterations if  $\mu$  is small.

For small values of  $\mu$ , (13) may thus be approximated by

$$\overline{\varepsilon}_{n+1} = (1 - \mu 2S_y) \overline{\varepsilon}_n + \mu^2 \widetilde{P}^2.$$
<sup>(19)</sup>

If  $1 - \mu 2S_y \neq 1$ , (19) can be rewritten as

$$\overline{\varepsilon}_n = \left(1 - \mu 2S_y\right)^n \overline{\varepsilon}_0 + \frac{\mu \widetilde{P}^2}{2S_y} \left[1 - \left(1 - \mu 2S_y\right)^n\right].$$
(20)

Under the condition

$$|1 - 2\mu S_y| < 1, (21)$$

the sequence (20) converges, and we get

$$\lim_{n \to \infty} \overline{\varepsilon}_n = \frac{\mu \widetilde{P}^2}{2S_y}.$$
(22)

Thus, the asymptotic deviation of  $||\mathbf{f}_n||^2$  from unity is proportional to the stepsize  $\mu$ . In general,  $\mu \ll 1$ , and so the algorithm (11), if stable, operates very close to unit filter norm.

Although the limit (22) is taken under the approximation of  $S_y$ and  $\tilde{P}^2$  being time invariant, it is suggested that (21) gives a sufficient condition for local convergence of  $\varepsilon_n$  over a limited span of iterations, over which time-invariance assumptions hold.

Condition (21) can be rewritten as

$$0 < \mu S_y < 1.$$
 (23)

Since  $\mu$  is positive by definition,  $S_y = E\{y_n^3\}$  is required to be positive, which is expected at convergence for a well-conditioned problem. Furthermore,  $S_y$  is expected to slowly increase over iterations as the algorithm converges in  $\mathbf{f}_n$ . The quantity  $\mu S_y$  could therefore be monitored during adaptation, and the stepsize decreased if necessary, in order to insure that (23) holds. This guidance on stepsize requires only a simple scalar multiplication and check.

Note that (23) only concerns stability in  $||\mathbf{f}_n||$ . A stepsize satisfying (23) is not guaranteed to give convergence to an  $\mathbf{f}_n$  maximizing the objective.



**Fig. 2.** Absolute deviation of  $\|\mathbf{f}_n\|$  from unity versus iteration number.



Fig. 3. Estimated skewness of filter output versus iteration.

# 5. EXPERIMENTAL RESULTS

In a numerical experiment, the algorithm (11) was used to implement an adaptive blind equalizer for a synthetic indoor Ultra-Wideband (UWB) communication channel with Impulse Radio signaling [9]. Such signals consist of pulse-position modulated impulses of extremely short duration, typically on the order of a nanosecond. Because of the large multipath spread of typical indoor UWB channels, intersymbol interference (ISI) is likely to occur at high data rates [10], [11]. Due to the impulsive nature of these signals, an adaptive blind equalizer based on third-order moments might be used to mitigate the effects of ISI.

The impulse response of an indoor UWB channel with a rich multipath spread of approximately 200 nanoseconds was synthesized with the aid of a recipe from [12], with additive white Gaussian noise at a signal-to-noise ratio per bit of 15dB. The Impulse Radio signals used binary orthogonal modulation at a bit rate of 10 Mbits/second and a sampling rate of 15 samples per nanosecond. Equalizers of order N = 400 were generated using (11) for three different stepsizes, each over 1000 adaptation iterations. All expectation operations in (11) were estimated using instantaneous values. Figures 2 and 3 show, respectively, the resulting absolute deviation of  $\|\mathbf{f}_n\|$  from unity and skewness versus iteration number. The plots show averaged results over 20 independent runs.

As seen in Figure 2, the deviation from unit norm at convergence increases with the stepsize, confirming the result (22) from Section 4. Figure 3 shows the convergence of the algorithm in terms of skewness. Note that a larger stepsize leads to faster convergence, but results in a smaller asymptotic skewness.

Experimental results also indicate that the stability condition (23) indeed can be monitored to indicate instability in  $\|\mathbf{f}_n\|$ . However, for stepsizes that give convergence in  $\mathbf{f}_n$  (as seen in Figure 3), (23) is typically satisfied by a large margin. Thus, for reasonable choices of  $\mu$ , the algorithm should be stable in  $\|\mathbf{f}_n\|$ .

# 6. CONCLUSION

A computationally efficient algorithm for norm-constrained gradient acsent has been studied for blind deconvolution of asymmetric source signals using a third-order moment based objective function. The results indicate that the algorithm maintains approximately unit filter norm for reasonable choices of adaptation stepsize. The condition on adaptation stepsize insuring a stable filter norm is trivial to calculate and verify. The small computational cost, involving only multiplications and additions, makes it well suited for on-line implementation on fixed-point digital signal processors.

# 7. REFERENCES

- Ralph A. Wiggins, "Minimum entropy deconvolution," *Geo-exploration*, no. 16, pp. 21–35, 1978.
- [2] David L. Donoho, "On minimum entropy deconvolution," in *Applied Time Series Analysis*, D. F. Findley, Ed. Academic Press, New York, 1981.
- [3] Patrik Pääjärvi and James P. LeBlanc, "Skewness maximization for impulsive sources in blind deconvolution," in *Proc. IEEE Nordic Signal Proc. Symp.*, Espoo, Finland, June 2004, pp. 304–307.
- [4] Patrik Pääjärvi and James P. LeBlanc, "On-line adaptive blind deconvolution based on third-order moments," *IEEE Signal Processing Lett.*, vol. 12, pp. 863–866, Dec. 2005.
- [5] Heinz Mathis and Scott C. Douglas, "Bussgang blind deconvolution for impulsive signals," *Proc. IEEE*, vol. 51, no. 7, pp. 1905–1915, July 2003.
- [6] Scott C. Douglas, Shun-ichi Amari, and S.-Y. Kung, "Gradient adaptation with unit-norm constraints," Tech. Rep. EE-99-003, Southern Methodist Univ., Dallas, Texas, 1999.
- [7] Scott C. Douglas, Shun-ichi Amari, and S.-Y. Kung, "On gradient adaptation with unit-norm constraints," *IEEE Trans. Signal Processing*, vol. 48, no. 6, pp. 1843–1847, June 2000.
- [8] Victor Solo and Xuan Kong, Adaptive Signal Processing Algorithms, Stability and Performance, Prentice-Hall, Englewood Cliffs, New Jersey, 1995.
- [9] Moe Z. Win and Robert A. Scholtz, "Ultra-wide bandwidth time-hopping spread-spectrum impulse radio for wireless multiple access communications," *IEEE Trans. Commun.*, vol. 48, no. 4, pp. 679–691, Apr. 2000.
- [10] Domenico Porcino and Walter Hirt, "Ultra-wideband radio technology: Potential and challenges ahead," *IEEE Commun. Mag.*, vol. 41, no. 7, pp. 66–74, July 2003.
- [11] A. G. Klein and C. R. Johnson, Jr., "MMSE decision feedback equalization of pulse position modulated signals," in *Communications, 2004 IEEE International Conference on*, Paris, France, June 2004, vol. 5, pp. 2648–2652.
- [12] Dajana Cassioli, Moe Z. Win, and Andreas F. Molisch, "The ultra-wide bandwidth indoor channel: From statistical model to simulations," *IEEE J. Select. Areas Commun.*, vol. 20, no. 6, pp. 1247–1257, Aug. 2002.