

A new method for kurtosis maximization and source separation

Marc Castella, Eric Moreau

► **To cite this version:**

Marc Castella, Eric Moreau. A new method for kurtosis maximization and source separation. ICASSP 2010: 35th International Conference on Acoustics, Speech, and Signal Processing , Mar 2010, Dallas, United States. IEEE, Proceedings ICASSP 2010: 35th International Conference on Acoustics, Speech, and Signal Processing pp.2670 - 2673, 2010, <10.1109/ICASSP.2010.5496250 >. <hal-01308032>

HAL Id: hal-01308032

<https://hal.archives-ouvertes.fr/hal-01308032>

Submitted on 27 Apr 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A NEW METHOD FOR KURTOSIS MAXIMIZATION AND SOURCE SEPARATION

Marc Castella

Institut TELECOM; TELECOM SudParis
Département CITI; UMR-CNRS 5157
9 rue Charles Fourier, 91011 Evry Cedex, France

Eric Moreau

University of Sud Toulon Var
ISITV, LSEET UMR-CNRS 6017
av. G. Pompidou, BP56
83162 La Valette du Var Cedex, France

ABSTRACT

This paper introduces a new method to maximize kurtosis-based contrast functions. Such contrast functions appear in the problem of blind source separation of convolutively mixed sources: the corresponding methods recover the sources one by one using a deflation approach. The proposed maximization algorithm is based on the particular nature of the criterion. The method is similar in spirit to a gradient ascent method, but differs in the fact that a “reference” contrast function is considered at each line search. The convergence of the method to a stationary point of the criterion can be proved. The theoretical result is illustrated by simulation.

Index Terms— Blind Source Separation, Contrast Function, Deflation, Higher-Order Statistics, Reference System, Optimization, Convergence

1. INTRODUCTION

The problem of blind source separation in a multi-input/multi-output convolutive context has found interesting solutions through the optimization of so-called contrast functions. Among the possible approaches, the source signals can be either separated simultaneously or extracted one by one by optimizing for each a multi-input/single-output separating criterion such as the constant modulus criterion or the kurtosis contrast. In this paper, we consider the latter approach, which is often combined with a deflation procedure to extract all the sources [1, 4, 5].

Contrast functions referred to as “reference-based” have been recently proposed [3]. They are particularly appealing because they are quadratic with respect to the searched parameters. They are an interesting alternative to the well-known kurtosis contrast functions [4], but they should be used cautiously, generally requiring a “fixed-point” like iteration to improve the separation quality [3].

In this paper, we propose a novel algorithm to maximize a kurtosis based contrast function. The proposed method is inspired from [2], but it is neither a gradient optimization of the kurtosis contrast function [4], nor a gradient optimization

of a “reference-based” contrast [2]. It is intermediate between the two methods and different from both.

2. MODEL AND ASSUMPTIONS

We consider an observed Q -dimensional discrete time signal $\mathbf{x}(n)$ (where $n \in \mathbb{Z}$ holds implicitly in the whole paper) which is given by the following convolutive mixing model:

$$\mathbf{x}(n) = \sum_{k \in \mathbb{Z}} \mathbf{M}(k)\mathbf{s}(n-k) \triangleq \{\mathbf{M}\}\mathbf{s}(n)$$

$\mathbf{M}(n)$ represents the $Q \times N$ matrix impulse response of the linear time invariant (LTI) mixing system and $\mathbf{s}(n)$ is a N -dimensional signal which components are referred to as the *sources*. The objective of source separation is to find an inverse separating LTI system. Our approach will be an iterative one where the sources are extracted one by one. Accordingly, we consider a $1 \times Q$ row filter, which impulse response will be denoted $\mathbf{w}(n)$. In case of successful separation, the output of the separator corresponds to one of the sources and is given by:

$$y(n) = \sum_{k \in \mathbb{Z}} \mathbf{w}(k)\mathbf{x}(n-k) \quad (1)$$

When the separation is performed using the observed signals $\mathbf{x}(n)$ only, the problem is referred to as the *blind source separation* (BSS) problem. To be able to solve the BSS problem, we introduce the following classical assumption:

- A1. The source signals $s_i(n), i \in \{1, \dots, N\}$ are stationary, zero-mean random processes with unit variance. Their fourth order cumulants exist and are non-zero. Finally, they are statistically mutually independent.

3. SEPARATION CRITERIA

It has been proved in [4, 5] that the criterion $\left| \frac{\text{Cum}\{y, y^*, y, y^*\}}{\mathbb{E}\{|y|^2\}^2} \right|$ is a contrast function under assumption A1: this means by definition that when maximized with respect to the separating

filter, this criterion leads to extraction of one source signal. In this paper, we propose a new method to maximize this kurtosis contrast function.

Our approach is tightly related to the concept of “reference-based” contrast functions [3]. We introduce a so-called “reference signal” which is given by the output of another $1 \times Q$ LTI separating filter denoted by $\{\mathbf{v}\}$.

$$z(n) = \sum_{k \in \mathbb{Z}} \mathbf{v}(k) \mathbf{x}(n-k) \triangleq \{\mathbf{v}\} \mathbf{x}(n) \quad (2)$$

In this paper, the reference signal is viewed as an efficient way to exploit the properties of the kurtosis contrast function. It is introduced for the purpose of facilitating the optimization: this is in contrast to former works, where the reference signal could be interpreted as an *a priori* information. Let us introduce the following criteria:

$$\mathcal{J}(\mathbf{w}) = \left| \frac{\text{Cum}\{y, y^*, y, y^*\}}{\mathbb{E}\{|y(n)|^2\}^2} \right|^2$$

$$\mathcal{I}(\mathbf{w}, \mathbf{v}) = \left| \frac{\text{Cum}\{y, y^*, z, z^*\}}{\mathbb{E}\{|y(n)|^2\} \mathbb{E}\{|z(n)|^2\}} \right|^2$$

The criterion \mathcal{J} is the well-known kurtosis contrast function [4, 5] whereas \mathcal{I} corresponds to so-called “reference contrasts” which have been recently introduced [3].

4. OPTIMIZATION METHOD

4.1. Algorithm

We now introduce our new algorithm for maximization of the kurtosis based contrast \mathcal{J} . First, we need to define a gradient operator ∇ and a partial gradient operators ∇_1 (respectively ∇_2) with respect to (w.r.t.) the first (respectively second) argument. More precisely, $\nabla \mathcal{J}(\mathbf{w})$ is the vector composed of all partial derivatives of $\mathcal{J}(\mathbf{w})$, whereas $\nabla_1 \mathcal{I}(\mathbf{w}, \mathbf{v})$ (respectively $\nabla_2 \mathcal{I}(\mathbf{w}, \mathbf{v})$) is the vector of partial derivatives of $\mathcal{I}(\mathbf{w}, \mathbf{v})$ w.r.t. \mathbf{w} (respectively \mathbf{v}). The algorithm is the following one:

Algorithm Alg.0

- Initialize \mathbf{v}_0 .
- For $k = 0, 1, 2, \dots$ repeat (M0):
 - (M0) * Set $\mathbf{d}_k = \nabla_1 \mathcal{I}(\mathbf{v}_k, \mathbf{v}_k)$,
 - * $\alpha_k = \arg \max_{\alpha} \mathcal{I}(\mathbf{v}_k + \alpha \mathbf{d}_k, \mathbf{v}_k)$ and set: $\mathbf{v}_{k+1} = \mathbf{v}_k + \alpha_k \mathbf{d}_k$.

The convergence property is cited next: it relies on the particular nature of the criteria \mathcal{J} and \mathcal{I} . Further improvements and practical consideration will be addressed in Section 5.

4.2. Convergence result

The proposed algorithm shows strong similarities with a steepest ascent algorithm: it can indeed be easily noticed that we have the symmetry $\mathcal{I}(\mathbf{w}, \mathbf{v}) = \mathcal{I}(\mathbf{v}, \mathbf{w})$ and that the criteria introduced in Section 3 are linked by the relation $\mathcal{J}(\mathbf{w}) = \mathcal{I}(\mathbf{w}, \mathbf{w})$. It then easily follows that:

$$\begin{aligned} \nabla_1 \mathcal{I}(\mathbf{w}, \mathbf{v}) &= \nabla_2 \mathcal{I}(\mathbf{v}, \mathbf{w}) \quad \text{and:} \\ \nabla \mathcal{J}(\mathbf{w}) &= 2 \nabla_1 \mathcal{I}(\mathbf{w}, \mathbf{w}) = 2 \nabla_2 \mathcal{I}(\mathbf{w}, \mathbf{w}) \end{aligned} \quad (3)$$

Hence, according to (3), the algorithm moves from one point to another moving along a gradient direction of the criterion \mathcal{J} . The noticeable difference is that during the one-dimensional optimization, the considered criterion is $\mathcal{I}(\mathbf{w}, \mathbf{v})$ with \mathbf{v} fixed instead of $\mathcal{J}(\mathbf{w})$. The convergence of the algorithm to a satisfying solution point hence requires justification. The following assumption is required:

- A2. The sources $s_i(n), i \in \{1, \dots, N\}$ are temporally independent and identically distributed (i.i.d.) Moreover, they have fourth-order cumulants which are all of the same sign.

Then we can state:

Proposition 1 Assume that the sequence $(\mathbf{v}_k)_{k \in \mathbb{N}}$ is obtained according to the algorithm **Alg.0** and that all $\mathbf{v}_k, k \in \mathbb{N}$ are contained in a compact set. Then, under assumption A2, any convergent subsequence of $(\mathbf{v}_k)_{k \in \mathbb{N}}$ converges to a point \mathbf{v}^* such that $\nabla \mathcal{J}(\mathbf{v}^*) = 0$.

The proof of Proposition 1 is based on the Zangwill’s convergence theorem [6]. Due to lack of space, it is skipped and postponed to a later publication.

4.3. Comments

Proposition 1 asserts the convergence of the algorithm to a stationary point of the criterion \mathcal{J} . However, similarly to the behavior of a steepest ascent algorithm, filters corresponding to minima or saddle-points of the criterion should never be obtained in practice. It appears in the proof that the algorithm yields a filter \mathbf{v}^* which maximizes $\mathcal{J}(\mathbf{v})$. Since it is known that any local maximum of the criterion corresponds to a separating filter [4], it follows that *the filters obtained with our algorithm are always separating ones in practice*. Unfortunately, in order to be able to state a general result, it is too strong to require that the algorithm yields a separating filter. The previous arguments however fully justify the importance of Proposition 1.

The previous result can also be understood from a different point of view: one can see that at each step, the algorithm **Alg.0** maximizes $\mathbf{w} \mapsto \mathcal{I}(\mathbf{w}, \mathbf{v}_k)$ along a gradient direction. This can be interpreted as a one-dimensional maximization of a “reference-based” contrast along the gradient direction. The proposed algorithm can thus be understood as an intermediate method lying between the following two methods:

- a gradient ascent on the kurtosis $\mathcal{J}(\mathbf{w})$: the difference is that a “reference-based” criterion is considered in **Alg.0** during each one-dimensional optimization.
- a gradient ascent on a “reference-based” contrast $\mathcal{I}(\mathbf{w}, \mathbf{v})$ with fixed “reference” \mathbf{v} [2]. In [2] is \mathbf{v} kept unchanged during the whole optimization, whereas here on the contrary, \mathbf{v} is updated after each one-dimensional optimization.

According to the latter point of view, the idea that the reference signal may contain *a priori* information on the separator re-enters, since the original “reference” is \mathbf{v}_0 and corresponds to the initialization point of algorithm **Alg.0**.

Note finally that for many sources, assumption A2 is quite reasonable: in particular, it holds systematically true in the context of digital communications. Since the validity of “reference-based” contrast functions has been proved in a more general context, the above explanations indicate however that the result of Proposition 1 is likely to remain true without assumption A2. This will be illustrated by computer simulations in Section 6.2.

5. IMPROVEMENTS AND IMPLEMENTATION

Due to the scaling indeterminacy, it is common in BSS to impose the constraint $\mathbb{E}\{|y(n)|^2\} = 1$. This can be done by introducing a re-normalizing step. Since the criteria in Section 3 are normalized so as to satisfy $\mathcal{I}(\lambda\mathbf{w}, \mathbf{v}) = \mathcal{I}(\mathbf{w}, \mathbf{v})$ and $\mathcal{J}(\lambda\mathbf{w}) = \mathcal{J}(\mathbf{w})$ for $\lambda \neq 0$, the re-normalization step introduced below does not alter the convergence property of the algorithm:

Algorithm Alg.1

- Initialize \mathbf{v}_0
- For $k = 0, 1, 2, \dots$ repeat (M0-R):
 - (M0) * Set $\mathbf{d}_k = \nabla_1 \mathcal{I}(\mathbf{v}_k, \mathbf{v}_k)$,
 - * $\alpha_k = \arg \max_{\alpha} \mathcal{I}(\mathbf{v}_k + \alpha \mathbf{d}_k, \mathbf{v}_k)$ and set: $\mathbf{v}_{k+1} = \mathbf{v}_k + \alpha_k \mathbf{d}_k$.
 - (R) Set: $\mathbf{v}_{k+1} = \frac{\mathbf{v}_{k+1}}{(\mathbb{E}\{\{\mathbf{v}_{k+1}\} \mathbf{x}(n)\})^{1/2}}$

Practically, we are only able to consider FIR separators of given length D . The conditions under which such a separator exists can be found in [1, 3]. The vectors of the impulse response can then be concatenated in the following $1 \times QD$ vector:

$$\underline{\mathbf{w}} \triangleq (\mathbf{w}(0) \quad \mathbf{w}(1) \quad \dots \quad \mathbf{w}(D-1))$$

We define $\underline{\mathbf{v}} \triangleq (\mathbf{v}(0), \mathbf{v}(1), \dots, \mathbf{v}(D-1))$ similarly and the $QD \times 1$ column vector:

$$\underline{\mathbf{x}}(n) \triangleq (\mathbf{x}(n)^T \quad \mathbf{x}(n-1)^T \quad \dots \quad \mathbf{x}(n-D+1)^T)^T.$$

Using these notations it is straightforward to see that the processed output and the “reference” signal defined in equations (1) and (2) can be written as:

$$y(n) = \underline{\mathbf{w}} \underline{\mathbf{x}}(n) \quad \text{and:} \quad z(n) = \underline{\mathbf{v}} \underline{\mathbf{x}}(n).$$

Now denote by $\mathbf{R} \triangleq \mathbb{E}\{\underline{\mathbf{x}}(n)\underline{\mathbf{x}}(n)^H\}$ the covariance matrix of $\underline{\mathbf{x}}(n)$ and define the matrix $\mathbf{C}(\underline{\mathbf{v}})$ component-wise by

$$(\mathbf{C}(\underline{\mathbf{v}}))_{i,j} = \text{Cum}\{\underline{x}_i(n), \underline{x}_j^*(n), z(n), z^*(n)\}.$$

where $\underline{x}_i(n), \underline{x}_j(n)$ are the components i and j of $\underline{\mathbf{x}}(n)$ respectively. Similarly to [2], we obtain:

$$\mathcal{I}(\underline{\mathbf{w}}, \underline{\mathbf{v}}) = |\tilde{\mathcal{I}}(\underline{\mathbf{w}}, \underline{\mathbf{v}})|^2 \quad \text{where:} \quad \tilde{\mathcal{I}}(\underline{\mathbf{w}}, \underline{\mathbf{v}}) = \frac{\underline{\mathbf{w}} \mathbf{C}(\underline{\mathbf{v}}) \underline{\mathbf{w}}^H}{\underline{\mathbf{w}} \mathbf{R} \underline{\mathbf{w}}^H}$$

Then, $\nabla_1 \mathcal{I}(\underline{\mathbf{v}}, \underline{\mathbf{w}})$ corresponds to the complex gradient vector $\underline{\mathbf{d}} = \frac{\partial \mathcal{I}}{\partial \underline{\mathbf{w}}^*}$ given below:

$$\frac{\partial \mathcal{I}}{\partial \underline{\mathbf{w}}^*} = \left(\frac{\partial \mathcal{I}}{\partial \underline{\mathbf{w}}} \right)^* = \left(2\tilde{\mathcal{I}}(\underline{\mathbf{w}}, \underline{\mathbf{v}}) \frac{\partial \tilde{\mathcal{I}}}{\partial \underline{\mathbf{w}}} \right)^* \quad \text{with:}$$

$$\frac{\partial \tilde{\mathcal{I}}}{\partial \underline{\mathbf{w}}} = \frac{\underline{\mathbf{w}}^* \mathbf{C}(\underline{\mathbf{v}})}{\underline{\mathbf{w}} \mathbf{R} \underline{\mathbf{w}}^H} - (\underline{\mathbf{w}}^* \mathbf{C}(\underline{\mathbf{v}}) \underline{\mathbf{w}}^H) \frac{\underline{\mathbf{w}}^* \mathbf{R}}{(\underline{\mathbf{w}} \mathbf{R} \underline{\mathbf{w}}^H)^2}$$

At the k th iteration of step (M0) of the algorithm, $\underline{\mathbf{d}}_k$ is given by the above equations where $\underline{\mathbf{v}}$ and $\underline{\mathbf{w}}$ are replaced by $\underline{\mathbf{v}}_k$. Finally, as explained in [2], the parameter α_k is a root of the polynomial $a_2 \alpha^2 + a_1 \alpha + a_0$ where:

$$\begin{aligned} a_2 &= \underline{\mathbf{d}}_k^H \mathbf{C}(\underline{\mathbf{v}}_k) \underline{\mathbf{d}}_k \Re[\underline{\mathbf{v}}_k \mathbf{R} \underline{\mathbf{d}}_k^H] - \Re[\underline{\mathbf{v}}_k \mathbf{C}(\underline{\mathbf{v}}_k) \underline{\mathbf{d}}_k] \underline{\mathbf{d}}_k^H \mathbf{R} \underline{\mathbf{d}}_k^H \\ a_1 &= \underline{\mathbf{d}}_k^H \mathbf{C}(\underline{\mathbf{v}}_k) \underline{\mathbf{d}}_k \underline{\mathbf{v}}_k \mathbf{R} \underline{\mathbf{v}}_k^H - \underline{\mathbf{v}}_k \mathbf{C}(\underline{\mathbf{v}}_k) \underline{\mathbf{v}}_k^H \underline{\mathbf{d}}_k \mathbf{R} \underline{\mathbf{d}}_k^H \\ a_0 &= \Re[\underline{\mathbf{v}}_k \mathbf{C}(\underline{\mathbf{v}}_k) \underline{\mathbf{d}}_k] \underline{\mathbf{v}}_k \mathbf{R} \underline{\mathbf{v}}_k^H - \underline{\mathbf{v}}_k \mathbf{C}(\underline{\mathbf{v}}_k) \underline{\mathbf{v}}_k^H \Re[\underline{\mathbf{v}}_k \mathbf{R} \underline{\mathbf{d}}_k^H] \end{aligned}$$

The value of α_k in (M0) is given by the root corresponding to the greatest value of $\mathcal{I}(\underline{\mathbf{v}}_k + \alpha \underline{\mathbf{d}}_k, \underline{\mathbf{v}}_k)$.

One can notice that, similarly to “reference-based” contrasts in [3], $\mathcal{I}(\underline{\mathbf{w}}, \underline{\mathbf{v}}_k)$ depends quadratically on $\underline{\mathbf{w}}$. It follows that the one-dimensional optimization step is easier to perform than in a classical gradient ascent method.

6. SIMULATIONS AND DISCUSSION

6.1. Validity and comparison

We first tested the validity of algorithm **Alg.1**. For different number of samples, $N = 3$ complex valued i.i.d. QAM4 sources have been generated taking values in $\{e^{i\pi/4}, e^{-i\pi/4}, e^{+i3\pi/4}, e^{-i3\pi/4}\}$ with equal probability 1/4. They have been mixed by mixing filters with randomly driven coefficients, length $L = 3$, and $Q = 4$ sensors. The results are reported in Table 1. A comparison has been made with former methods, in particular a gradient maximization of the same criterion \mathcal{J} and the results given by a reference based contrast function [2]. Both the average and median values of the mean square

error (MSE) over 1000 Monte-Carlo realizations are reported in Table 1. It can be observed that the result given by a gradient maximization and by algorithm **Alg.1** are similar, which illustrates the validity of our method.

	separation method	Number of samples		
		1000	5000	10000
mean	kurtosis	0.0031	0.0030	0.0028
	ref. contrast	0.3752	0.1286	0.0842
	Alg.1	0.0032	0.0030	0.0028
median	kurtosis	0.0008	0.0007	0.0007
	ref. contrast	0.2926	0.0603	0.0316
	Alg.1	0.0009	0.0007	0.0007

Table 1. Average (1000 realizations) MSE for different contrast function and optimization methods. (QAM4, $N=3$, $Q=4$, $L=3$)

6.2. Influence of assumption A2

Although we are unable so far to prove Proposition 1 when Assumption A2 is relaxed, we can expect that it remains valid in a much broader context as explained in Section 4.3. The following situations have been tested:

Different signs of the sources' cumulants: We generated two zero-mean, unit variance, uniformly distributed, i.i.d. sources (cumulant value -1.2) and one unit-variance Laplace i.i.d. source (cumulant value $+3$). The successive sources have been retrieved using a deflation approach. The separation results and average MSE values are gathered in Table 2. The values indicate that the method seems still valid in this case.

Non i.i.d. sources: We considered Continuous Phase Modulation (CPM) source signals, which are non i.i.d. and of particular interest in a communication application. In addition, the general scalar filtering ambiguity that generally remains when separating non i.i.d. sources has been characterized in the case of a kurtosis-based contrast function [1]. It could hence be promising to test our optimization method with CPM sources. A typical source separation result is given on Figure 1: it illustrates that our method seems to perform similarly to the kurtosis based method [1]. Note that the separation quality of the second source is much worse than the first one: this is a classical drawback of deflation methods. All previously proposed solutions [1] can be used to avoid this problem.

7. REFERENCES

[1] M. Castella, P. Bianchi, A. Chevreuil, and J.-C. Pesquet. A blind source separation framework for detecting CPM sources mixed by a convolutive MIMO filter. *Signal Processing*, 86(8):1950–1967, August 2006.

	separation method	Source number		
		1st	2nd	3rd
mean	kurtosis	0.0265	0.0330	0.0604
	ref. contrast	0.0856	0.1133	0.1779
	Alg.1	0.0220	0.0431	0.1009
median	kurtosis	0.0248	0.0155	0.0203
	ref. contrast	0.0674	0.0887	0.1284
	Alg.1	0.0199	0.0174	0.0468

Table 2. Average and median (1000 realizations) MSE for the separation of sources not satisfying A.2 (2 uniform sources and 1 doubly exponential, $N=3$, $Q=4$, $L=3$, 5000 samples)

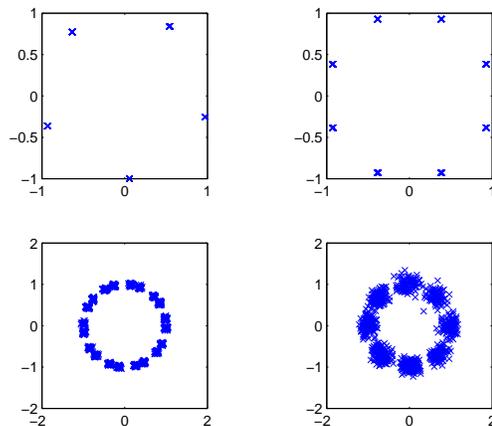


Fig. 1. Typical separation result of two CPM sources ($N=2$, $Q=3$, $L=5$). Original sources (top) and separated sources (bottom).

[2] M. Castella and E. Moreau. A new optimization method for reference-based quadratic contrast functions in a deflation scenario. In *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3161–3164, Taipei, Taiwan, R.O.C., April 2009.

[3] M. Castella, S. Rhioui, E. Moreau, and J.-C. Pesquet. Quadratic higher-order criteria for iterative blind separation of a MIMO convolutive mixture of sources. *IEEE Trans. Signal Processing*, 55(1):218–232, January 2007.

[4] C. Simon, P. Loubaton, and C. Jutten. Separation of a class of convolutive mixtures: a contrast function approach. *Signal Processing*, (81):883–887, 2001.

[5] J. K. Tugnait. Identification and deconvolution of multi-channel linear non-gaussian processes using higher order statistics and inverse filter criteria. *IEEE Trans. Signal Processing*, 45(3):658–672, March 1997.

[6] W. I. Zangwill. *Nonlinear Programming: A Unified Approach*. Prentice-Hall, Inc. Englewood Cliffs, N.J., 1969.