

COMPARISON OF CONVEX COMBINATION AND AFFINE COMBINATION OF ADAPTIVE FILTERS

Alper T. Erdogan, Suleyman S. Kozat

Andrew C. Singer

Koc University
Istanbul, Turkey

University of Illinois at Urbana-Champaign
IL, USA

ABSTRACT

In the area of combination of adaptive filters, two main approaches, namely convex and affine combinations have been introduced. In this article, the relation between these two approaches is investigated. First, the problem of obtaining optimal convex combination coefficients is formulated as the projection of the optimal affine combination weights to the unit simplex in a weighted inner product space. Based on this formulation the closed form expressions for optimal combination weights and target MSE levels are obtained for two and three branch cases.

Index Terms— adaptive filtering, combination of adaptive filters, projection to unit simplex

1. INTRODUCTION

The combination of multiple adaptive branches designated for the same task has been proposed as a robust solution to model (order, algorithm, parameter) selection problem in adaptive filtering. Earlier works in this field concentrated on the case where the convex combination scheme is applied at the outputs [1, 2]. Later this approach has been extended by relaxing the convex combination requirement to the affine combination requirement [3]. In this article, we provide an analysis to compare target MSE levels of these two adaptive combination methods.

2. AFFINE AND CONVEX COMBINATION OF ADAPTIVE FILTERS

2.1. Data Model

We assume that there are m adaptive filtering branches with (zero mean) outputs $y_k(t)$, $k = 1, \dots, m$ to be combined. Defining

$$\mathbf{y}(t) = [y_1(t) \quad y_2(t) \quad \dots \quad y_m(t)]^T$$

This work is supported in part by TUBITAK Career Awards, Contract Nos:104E073 and 108E195, and Turkish Academy of Sciences GEBIP Program.

as the input vector to the combination block and assuming that $d(t)$ is the desired signal, with zero mean and variance σ_d^2 , we can define the limiting covariances

$$\begin{aligned} \mathbf{R} &\triangleq \lim_{t \rightarrow \infty} E(\mathbf{y}(t)\mathbf{y}^T(t)) \\ \mathbf{p} &\triangleq \lim_{t \rightarrow \infty} E(\mathbf{y}(t)d(t)). \end{aligned}$$

We formulate the optimal combination problem as finding the combination weight vector $\boldsymbol{\omega}$ that will minimize the MSE between the desired signal and the combined output.

2.2. Affine Combination Problem

We can pose the problem of finding the best weights to obtain affine combination of multiple adaptive filtering algorithms as the convex optimization problem (*Problem 1*)

$$\begin{aligned} \text{minimize} \quad & J^a(\boldsymbol{\omega}^a) = \sigma_d^2 - \mathbf{p}^T \mathbf{R}^{-1} \mathbf{p} + (\boldsymbol{\omega}^a - \boldsymbol{\omega}_o)^T \mathbf{R} (\boldsymbol{\omega}^a - \boldsymbol{\omega}_o) \\ \text{subject to} \quad & \mathbf{1}^T \boldsymbol{\omega}^a = 1, \end{aligned}$$

where $\boldsymbol{\omega}_o \triangleq \mathbf{R}^{-1} \mathbf{p}$ and $\mathbf{1}$ is a vector of all ones.

We can write a closed form solution to this linearly constrained quadratic minimization problem, through use of Lagrangian formulation,

$$\boldsymbol{\omega}_o^a \triangleq \boldsymbol{\omega}_o + \frac{(1 - \mathbf{1}^T \boldsymbol{\omega}_o)}{\mathbf{1}^T \mathbf{R}^{-1} \mathbf{1}} \mathbf{R}^{-1} \mathbf{1}, \quad (1)$$

and the corresponding MSE is given by

$$J_o^a \triangleq \sigma_d^2 - \mathbf{p}^T \mathbf{R}^{-1} \mathbf{p} + \frac{(1 - \mathbf{1}^T \boldsymbol{\omega}_o)^2}{\mathbf{1}^T \mathbf{R}^{-1} \mathbf{1}}. \quad (2)$$

2.3. Convex Combination Problem

The convex combination case can also be posed as the convex quadratic minimization problem with linear constraints (*Problem 2*):

$$\begin{aligned} \text{minimize} \quad & J^c(\boldsymbol{\omega}^c) = \sigma_d^2 - \mathbf{p}^T \mathbf{R}^{-1} \mathbf{p} + (\boldsymbol{\omega}^c - \boldsymbol{\omega}_o)^T \mathbf{R} (\boldsymbol{\omega}^c - \boldsymbol{\omega}_o) \\ \text{subject to} \quad & \mathbf{1}^T \boldsymbol{\omega}^c = 1 \\ & \boldsymbol{\omega}^c \geq \mathbf{0}. \end{aligned}$$

The corresponding constraint set $\Delta = \{\omega | \mathbf{1}^T \omega = 1, \omega \geq \mathbf{0}\}$ is referred as the unit or standard simplex in the literature. The optimization problem corresponding to the convex combination doesn't have a nice closed form solution in general. However, we can still provide some analytical characterization for the solution of the optimization problem corresponding to the selection of the best convex combination vector.

We first note that we can rewrite the cost function as

$$\begin{aligned}
J^c(\omega^c) &= \sigma_d^2 - \mathbf{p}^T \mathbf{R}^{-1} \mathbf{p} + (\omega^c - \omega_o)^T \mathbf{R} (\omega^c - \omega_o) \quad (3) \\
&= \sigma_d^2 - \mathbf{p}^T \mathbf{R}^{-1} \mathbf{p} \\
&+ ((\omega^c - \omega_o^a) - (\omega_o^a - \omega_o))^T \mathbf{R} ((\omega^c - \omega_o^a) - (\omega_o^a - \omega_o)) \\
&= \sigma_d^2 - \mathbf{p}^T \mathbf{R}^{-1} \mathbf{p} \\
&+ (\omega_o^a - \omega_o)^T \mathbf{R} (\omega_o^a - \omega_o) \\
&+ (\omega^c - \omega_o^a)^T \mathbf{R} (\omega^c - \omega_o^a) \\
&- 2(\omega^c - \omega_o^a)^T \mathbf{R} (\omega_o^a - \omega_o) \\
&= J_o^a + (\omega^c - \omega_o^a)^T \mathbf{R} (\omega^c - \omega_o^a) \\
&- 2 \frac{(1 - \mathbf{1}^T \omega_o)}{\mathbf{1}^T \mathbf{R}^{-1} \mathbf{1}} (\omega^c - \omega_o^a)^T \mathbf{1} \quad (4) \\
&= J_o^a + (\omega^c - \omega_o^a)^T \mathbf{R} (\omega^c - \omega_o^a), \quad (5)
\end{aligned}$$

where we used the fact that over the unit simplex the last term in (4) is equal to zero and use (1). Therefore, ignoring the constant term in the cost function we can pose the problem of finding ω_o^c as the problem of projecting ω_o^a to the unit simplex Δ_m with respect to the \mathbf{R} -weighted 2-norm (where Δ_m is the m dimensional unit simplex). In other words, the problem of finding the best ω^c can be posed as (*Problem 3*)

$$\begin{aligned}
&\text{minimize} && \|\omega_o^a - \omega^c\|_{\mathbf{R}}^2 \\
&\text{subject to} && \omega^c \in \Delta_m.
\end{aligned}$$

In the special case where $\mathbf{R} = \mathbf{I}$, i.e., when we use the standard Euclidian distance, the corresponding problem is relatively simpler, which can be solved via an algorithm with finite steps[4]. However, for a general \mathbf{R} matrix, the optimal point characterization is more involved.

2.4. Two Branches

When we have two adaptive filtering branches to be combined (i.e. $m = 2$ case), the unit simplex Δ_2 is simply the line segment between $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$. In this case, we have two major cases to consider:

- Case $\omega_o^a \in \Delta_2$: In this case, we simply have $\omega_o^c = \omega_o^a$ and $J_o^a = J_o^c$. This case is illustrated in Figure 1 (a).
- Case $\omega_o^a \notin \Delta_2$: This case refers to one of the components of ω_o^a is strictly negative and the other one is positive. (We consider the first component to be negative, without the loss of generality). Due to $\mathbf{1}^T \omega_o^a =$

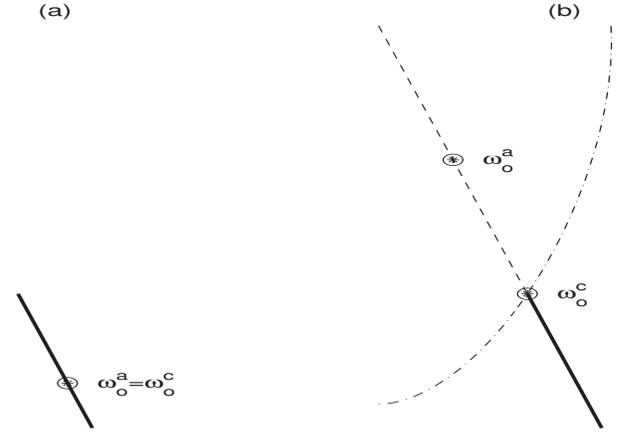


Fig. 1. Combination of 2 Adaptive Stages: Affine vs. Convex Combination

$\mathbf{1}^T \omega^c = 1$ constraint, we can write

$$\omega^c = \omega_o^a + \alpha_c \begin{bmatrix} 1 & -1 \end{bmatrix}^T. \quad (6)$$

Therefore, the cost function would be equivalent to

$$\|\alpha_c \begin{bmatrix} -1 \\ 1 \end{bmatrix}\|_{\mathbf{R}}^2 = \alpha_c^2 \begin{bmatrix} -1 & 1 \end{bmatrix} \mathbf{R} \begin{bmatrix} -1 \\ 1 \end{bmatrix},$$

which implies that the cost increases when the magnitude of α_c increases. Therefore, the smallest value which makes $\omega^c = \omega_o^a + \alpha_c \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ feasible (i.e., in Δ_2) would be the optimal choice for α_c which is equivalent to $-\omega_o^a(1)$. This would be equivalent to the choice $\omega_o^c = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, i.e. a corner point of Δ_2 . This case is illustrated in Figure 1 (b), where the level set of the cost function (i.e. the ellipse segment) corresponding to the minimum cost value is also shown. As a result the cost increase for the convex combination relative to the affine combination case can be written as

$$J_o^c - J_o^a = \begin{cases} \zeta^a \omega_o^a(1)^2 & \omega_o^a(1) < 0 \\ \zeta^a \omega_o^a(2)^2 & \omega_o^a(2) < 0 \\ 0 & \text{otherwise} \end{cases}, \quad (7)$$

where $\zeta^a = \begin{bmatrix} -1 \\ 1 \end{bmatrix}^T \mathbf{R} \begin{bmatrix} -1 \\ 1 \end{bmatrix} = R_{11} + R_{22} - 2R_{12}$. Here we can write a closed form expression for the optimal affine combination weights as

$$\omega_o^a = \frac{1}{\zeta^a} \begin{bmatrix} (R_{22} - R_{12})(1 - \omega_o(2)) + (R_{11} - R_{12})\omega_o(1) \\ (R_{11} - R_{12})(1 - \omega_o(1)) + (R_{22} - R_{12})\omega_o(2) \end{bmatrix}.$$

2.5. Three Branches

In the Three Branch Case, the constraint set Δ_3 is a triangle, located inside the 2-dimensional plane $A = \{\omega \mid \mathbf{1}^T \omega =$

$1, \omega \in \mathfrak{R}^3$ corresponding to the affine combination constraint. We can represent the unit simplex Δ_3 as the intersection of A with halfspaces

$$H_i = \{\omega \mid \mathbf{e}_i^T \omega \geq 0\} \quad i = 1, 2, 3. \quad (8)$$

Since the optimal affine combination vector ω_o^a is located in the same plane, the corresponding projection problem can be analyzed over the same plane. If $\omega_o^a \in \Delta_3$ then, clearly, $\omega_o^c = \omega_o^a$. However, if one or more of the entries are negative then $\omega_o^a \notin \Delta_3$ needs to be projected to Δ_3 . For this purpose, we introduce a mapping \mathcal{T} which maps the affine space A to the two dimensional vector space \mathfrak{R}^2 :

$$\beta^a = \mathcal{T}(\omega^a) = \mathbf{Q}\omega^a, \quad \omega^a \in A \quad (9)$$

where

$$\mathbf{Q} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}. \quad (10)$$

The inverse of this mapping is given by

$$\omega^a = \mathbf{V}\beta^a + \mathbf{e}_3, \quad \beta^a \in \mathfrak{R}^2, \quad (11)$$

where

$$\mathbf{V} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -1 & -1 \end{bmatrix}. \quad (12)$$

If we assume that $\omega_o^a(1) < 0$, without loss of generality, we can start the Dykstra's Algorithm [5] by projecting ω_o^a to $H_1 \cap A$. This projection operation is performed in the inner product space equipped with

$$\langle \mathbf{x}, \mathbf{y} \rangle_{\mathbf{R}} = \mathbf{y}^T \mathbf{R} \mathbf{x}, \quad \mathbf{x}, \mathbf{y} \in \mathfrak{R}^3, \quad (13)$$

as the inner product, which induces the weighted norm $\|\cdot\|_{\mathbf{R}}$. By invoking the projection theorem which asserts that the projection error should be orthogonal to the target space, we obtain the projection rule as

$$\mathcal{P}_{H_1 \cap A}(\omega_o^a) = \omega_o^a - \frac{\omega_o^a(1)}{\mathbf{e}_1^T \mathbf{R} \mathbf{V}^{-1} \mathbf{e}_1} \mathbf{V} \mathbf{R} \mathbf{V}^{-1} \mathbf{e}_1, \quad (14)$$

where $\mathbf{R}_{\mathbf{V}} = \mathbf{V}^T \mathbf{R} \mathbf{V}$. We can write the expression above more explicitly, after some algebraic simplifications, in terms of elements of \mathbf{R} as

$$\mathcal{P}_{H_1 \cap A}(\omega_o^a) = \omega_o^a - \omega_o^a(1) \mathbf{q}_1 \quad (15)$$

where

$$\mathbf{q}_1 = \begin{bmatrix} 1 & \frac{R_{13}+R_{23}-R_{33}-R_{12}}{R_{22}+R_{33}-2R_{23}} & \frac{R_{12}+R_{23}-R_{22}-R_{13}}{R_{22}+R_{33}-2R_{23}} \end{bmatrix}^T. \quad (16)$$

If the above projection maps ω_o^a to a point on Δ_3 , i.e., the resulting projection point has non-negative elements, then Dykstra's algorithm terminates at that point. Therefore, the resulting point would also be the projection of the set Δ_3 . We can

characterize the points $\omega^a \in A$ that would be projected to the side $H_1 \cap \Delta_3$ as

$$\mathbf{e}_1^T \omega^a \leq 0, \quad \mathbf{e}_2^T \mathcal{P}_{H_1 \cap A}(\omega^a) \geq 0, \quad \mathbf{e}_3^T \mathcal{P}_{H_1 \cap A}(\omega^a) \geq 0.$$

This set can be more explicitly defined as

$$\Upsilon_1 = \{\omega^a \mid -\mathbf{e}_1^T \omega^a \geq 0, \mathbf{h}_{12}^T \omega^a \geq 0, \mathbf{h}_{13}^T \omega^a \geq 0\},$$

where

$$\begin{aligned} \mathbf{h}_{12}^T &= [R_{33} + R_{12} - R_{13} - R_{23} \quad R_{22} + R_{33} - 2R_{23} \quad 0] \\ \mathbf{h}_{13}^T &= [R_{22} + R_{13} - R_{23} - R_{12} \quad 0 \quad R_{22} + R_{33} - 2R_{23}] \end{aligned}$$

By following a similar procedure, Υ_2 and Υ_3 , which are polyhedral regions that are projected to the line segments $H_2 \cap \Delta_3$ and $H_3 \cap \Delta_3$, respectively, can be obtained as follows:

$$\Upsilon_i = \{\omega^a \mid -\mathbf{e}_i^T \omega^a \geq 0, \mathbf{h}_{ij}^T \omega^a \geq 0 \quad 1 \leq j \neq i \leq 3\}, \quad i = 2, 3$$

where

$$\begin{aligned} \mathbf{h}_{21}^T &= [R_{11} + R_{33} - 2R_{13} \quad R_{33} + R_{12} - R_{13} - R_{23} \quad 0] \\ \mathbf{h}_{23}^T &= [0 \quad R_{11} + R_{23} - R_{12} - R_{13} \quad R_{11} + R_{33} - 2R_{13}] \\ \mathbf{h}_{31}^T &= [R_{11} + R_{22} - 2R_{12} \quad 0 \quad R_{33} + R_{13} - R_{12} - R_{23}] \\ \mathbf{h}_{32}^T &= [0 \quad R_{11} + R_{22} - 2R_{12} \quad R_{11} + R_{23} - R_{12} - R_{13}] \end{aligned}$$

We also define the remaining neighboring regions as

$$\begin{aligned} \Upsilon_{12} &= \{\omega^a \mid [\mathbf{h}_{12} \quad \mathbf{h}_{21}]^T \omega^a < 0\} \\ \Upsilon_{13} &= \{\omega^a \mid [\mathbf{h}_{13} \quad \mathbf{h}_{31}]^T \omega^a < 0\} \\ \Upsilon_{23} &= \{\omega^a \mid [\mathbf{h}_{23} \quad \mathbf{h}_{32}]^T \omega^a < 0\}. \end{aligned}$$

We'll now show that if ω_o^a is in Υ_{ij} then $\omega_o^c = \mathbf{e}_k$ where $k \neq i, j$ and $1 \leq k \leq 3$. In order to prove this fact, we will use KKT optimality conditions for *Problem 3*. We first write the corresponding Lagrangian function

$$L(\omega^c, \lambda, \mu) = \|\omega_o^a - \omega^c\|_{\mathbf{R}}^2 - \lambda^T \omega^c + \mu(\mathbf{1}^T \omega^c - 1).$$

According to KKT conditions, the primal vector ω_o^c and dual variables λ_o and μ_o correspond to the optimal point if and only the following equations hold:

- (i) $\omega_o^c \geq 0, \mathbf{1}^T \omega_o^c = 1$
- (ii) $\lambda_o(i) \omega_o^c(i) = 0, i = 1, \dots, 3$
- (iii) $\lambda_o \geq 0$
- (iv) $\nabla L(\omega_o^c, \lambda_o, \mu_o) = 0.$

The last equality implies that

$$(iv)' \quad \lambda_o = 2\mathbf{R}(\omega_o^c - \omega_o^a) + \mu_o \mathbf{1}.$$

If we were to derive conditions on \mathbf{R} and ω_o^a for $\omega_o^c = \mathbf{e}_3$:

For this case we can write

$$\begin{aligned} \omega_o^c - \omega_o^a &= \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} - \begin{bmatrix} \omega_o^a(1) \\ \omega_o^a(2) \\ (1 - \omega_o^a(1) - \omega_o^a(2)) \end{bmatrix} \\ &= \begin{bmatrix} -\omega_o^a(1) \\ -\omega_o^a(2) \\ \omega_o^a(1) + \omega_o^a(2) \end{bmatrix}. \end{aligned}$$

The fact that $\omega_o^c(3) = 1 > 0$ implies $\lambda_o(3) = 0$ due to (ii). Therefore, using this fact, from the third row of (iv)' we find

$$\mu_o = 2((R_{13} - R_{33})\omega_o^a(1) + (R_{23} - R_{33})\omega_o^a(2)). \quad (17)$$

Therefore, from the condition $\lambda_o(1) \geq 0$ and the first row of (iv)', we obtain

$$-\mathbf{h}_{21}^T \omega_o^a \geq 0 \quad (18)$$

and similarly, from the second row of (iv)' and the condition $\lambda_o(2) \geq 0$ we have

$$-\mathbf{h}_{12}^T \omega_o^a \geq 0. \quad (19)$$

Therefore, $\omega_o^c = \mathbf{e}_3$ if and only if $\omega_o^a \in \Upsilon_{12}$. By following a similar procedure, one can show that $\omega_o^c = \mathbf{e}_2$ for $\omega_o^a \in \Upsilon_{13}$ and $\omega_o^c = \mathbf{e}_1$ for $\omega_o^a \in \Upsilon_{23}$.

As a result, we can summarize the formulation of ω_o^c in terms of ω_o^a as

$$\omega_o^c = \begin{cases} \omega_o^a & \omega_o^a \geq 0 \\ \omega_o^a - \omega_o^a(i)\mathbf{q}_i & \omega_o^a \in \Upsilon_i \\ \mathbf{e}_1 & \omega_o^a \in \Upsilon_{23} \\ \mathbf{e}_2 & \omega_o^a \in \Upsilon_{13} \\ \mathbf{e}_3 & \omega_o^a \in \Upsilon_{12} \end{cases},$$

where \mathbf{q}_1 is as defined in (16) and

$$\begin{aligned} \mathbf{q}_2 &= \begin{bmatrix} \frac{R_{23}+R_{13}-R_{33}-R_{12}}{R_{11}+R_{23}-2R_{13}} & 1 & \frac{R_{21}+R_{13}-R_{11}-R_{23}}{R_{11}+R_{33}-2R_{13}} \end{bmatrix}^T \\ \mathbf{q}_3 &= \begin{bmatrix} \frac{R_{23}+R_{12}-R_{22}-R_{13}}{R_{11}+R_{22}-2R_{12}} & \frac{R_{13}+R_{12}-R_{11}-R_{23}}{R_{11}+R_{22}-2R_{12}} & 1 \end{bmatrix}^T. \end{aligned}$$

The corresponding relative MSE misadjustment is given by

$$J_o^c - J_o^a = \begin{cases} 0 & \omega_o^a \geq 0 \\ |\omega_o^a(i)|^2 \mathbf{q}_i^T \mathbf{R} \mathbf{q}_i & \omega_o^a \in \Upsilon_i \\ \|\omega_o^a - \mathbf{e}_1\|_{\mathbf{R}}^2 & \omega_o^a \in \Upsilon_{23} \\ \|\omega_o^a - \mathbf{e}_2\|_{\mathbf{R}}^2 & \omega_o^a \in \Upsilon_{13} \\ \|\omega_o^a - \mathbf{e}_3\|_{\mathbf{R}}^2 & \omega_o^a \in \Upsilon_{12} \end{cases}. \quad (20)$$

Figure 2 illustrates the projection problem corresponding to obtaining optimal convex combination coefficients from affine combination coefficients. On the plane corresponding to the affine combination constraint, the unit simplex corresponding to the convex combination weights and the polyhedral partitions that specify whether the affine combination weights are mapped to the sides or the corners of the unit simplex are shown. A case where ω_o^a is in Υ_{23} and mapped to \mathbf{e}_1 is shown, where the level set of the weighted distance cost function is also drawn to demonstrate the nature of projection.

2.6. General Case

The analysis above converted the quadratic programming with unit simplex constraint into a simple problem of polyhedral inclusion. The polyhedral partitioning of the affine constraint set approach could be extended to the case of higher dimensional spaces, where the expressions get more involved.

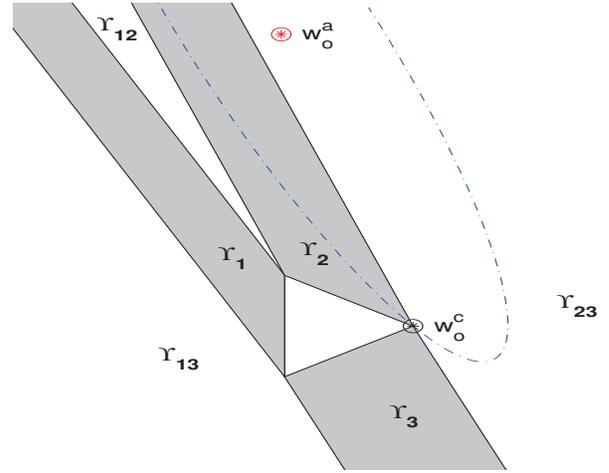


Fig. 2. Combination of Three Adaptive Branches: Affine vs. Convex Combination.

3. REFERENCES

- [1] A.C. Singer and M. Feder, "Universal linear prediction by model order weighting," *IEEE Trans. on Signal Processing*, vol. 47, no. 10, pp. 2685–2699, October 1999.
- [2] J. Arenas-Garcia, A.R. Figueiras-Vidal, and A.H. Sayed, "Mean square performance of a convex combination of two adaptive filters," *IEEE Trans. on Signal Processing*, vol. 54, no. 3, pp. 1078–1090, March 2006.
- [3] J.Y. Tournet, N. J. Bershad, J.C.M. Bermudez, "An affine combination of two lms adaptive filters- transient mean square analysis," *IEEE Trans. on Signal Processing*, vol. 56, no. 5, May 2008.
- [4] H. J. H. Tuentler, "The minimum l_2 distance projection onto the canonical simplex: A simple algorithm," *Algo Research Quarterly*, vol. 4, pp. 53–55, December 2001.
- [5] J.P. Boyle and R. L. Dykstra, "A method for finding projections onto the intersection of convex sets in Hilbert space," *Advances in Order Restricted Statistical Inference*, vol. 37, pp. 28–47, 1986.