World Academy of Science, Engineering and Technology
International Journal of Physical and Mathematical Sciences
Vol:1, No:6, 2007

# A Simple Adaptive Algorithm for Norm-Constrained Optimization

Hyun-Chool Shin *Member, IEEE*



***Abstract***— In this paper we propose a simple adaptive algorithm iteratively solving the unit-norm constrained optimization problem. Instead of conventional parameter norm based normalization, the proposed algorithm incorporates scalar normalization which is computationally much simpler. The analysis of stationary point is presented to show that the proposed algorithm indeed solves the constrained optimization problem. The simulation results illustrate that the proposed algorithm performs as good as conventional ones while being computationally simpler.

***Keywords***— constrained optimization, unit-norm, LMS, principle component analysis.

## I. INTRODUCTION

Constrained optimization is the minimization of an objective function subjected to constraints on the possible values of the independent parameters. A typical problem arising in signal processing for the parameter vector $\mathbf{w} = [\ w(0) \quad w(1) \quad \cdots \quad w(K-1)\ ]^T$ is to minimize a quadratic cost function of an error

$$e_i = (d_i - \mathbf{w}^T\mathbf{x}_i)^2 \tag{1}$$

where $d(i)$ is the desired data and $\mathbf{x}_i$ denotes the $(K\times 1)$ input vector, $\mathbf{x} = [\ x_i(0) \quad x_i(1) \quad \cdots \quad x_i(K-1)\ ]^T$. A general form of the constrained optimization is

$$\min_{\mathbf{w}} J(\mathbf{w}) = f(e_i) \tag{2}$$

subject to

$$\|\mathbf{w}\|^2 = \sum_{k=0}^{K-1} |w(k)|^2 = 1.$$

This is one of the basic and fundamental problems in communications, controls and adaptive signal processing [1]–[7].

The well-known stochastic gradient algorithm for minimizing (2) has the following form

$$\begin{aligned}\tilde{\mathbf{w}}_i &= \mathbf{w}_{i-1} - \mu\frac{\partial J(\mathbf{w})}{\partial \mathbf{w}_{i-1}}\\ &= \mathbf{w}_{i-1} - \mu\frac{\partial f(e_i)}{\partial e_i}\frac{\partial e_i}{\partial \mathbf{w}_{i-1}}\\ &= \mathbf{w}_{i-1} + \mu\phi_i\mathbf{x}_i \end{aligned} \tag{3}$$

$$\mathbf{w}_i = \frac{\tilde{\mathbf{w}}_i}{\|\tilde{\mathbf{w}}_i\|}, \tag{4}$$

where $\mathbf{w}_i$ is an estimate at iteration $i$, $\mu$ is the step-size, $\phi_i = \frac{\partial f(e_i)}{\partial e_i}$ and $\frac{\partial e_i}{\partial \mathbf{w}_{i-1}} = -\mathbf{x}_i$. The adaptive algorithm is realized in two steps. The first step is the same as the conventional stochastic gradient adaptive algorithm which minimizes $J(\mathbf{w})$ itself without any constraint. Then, in the second step, the constraint is separately kept by normalizing $\tilde{\mathbf{w}}_i$ by $\|\tilde{\mathbf{w}}_i\|$ so that the algorithm satisfies $\|\mathbf{w}_i\| = 1$ at every iteration.

This work was supported by the Soongsil University Research Fund.
H.-C. Shin is with School of Electronic Engineering, College of Information Technology, Soongsil Univ., Seoul, Republic of Korea (shinhc@ssu.ac.kr).

## II. THE PROPOSED ADAPTIVE ALGORITHM WITH SCALAR NORMALIZATION

Here we propose a simpler adaptive algorithm solving unit-norm constrained optimization problem. Instead of conventional parameter norm based normalization in (3)-(4), the proposed method incorporates scalar normalization:

$$\begin{aligned}\tilde{\mathbf{w}}_i &= \mathbf{w}_{i-1} + \mu\phi_i\mathbf{x}_i, \text{ and}\\ \mathbf{w}_i &= \frac{\tilde{\mathbf{w}}_i}{1+\mu\phi_i y_i},\end{aligned} \tag{5}$$

where $y_i = \mathbf{w}_i^T\mathbf{x}_i$.

Now to verify that the proposed method in (5) solves the unit-norm constrained optimization problem in (2), we provide the steady-state analysis. The equation in (5) can be written in

$$\mathbf{w}_i = \frac{\mathbf{w}_{i-1} + \mu\phi_i\mathbf{x}_i}{1+\mu\phi_i y_i}. \tag{6}$$

Subtracting $\mathbf{w}_{i-1}$ from both sides of (6), we get

$$\begin{aligned}\mathbf{w}_i - \mathbf{w}_{i-1} &= \frac{\mathbf{w}_{i-1} + \mu\phi_i\mathbf{x}_i}{1+\mu\phi_i y_i} - \mathbf{w}_{i-1}\\ &= \frac{\mu\phi_i(\mathbf{x}_i - y_i\mathbf{w}_{i-1})}{1+\mu\phi_i y_i}\end{aligned} \tag{7}$$

Since the denominator in the left side of (7) is

$$\frac{1}{1+\mu\phi_i y_i} = 1 + O(\mu), \tag{8}$$

the equation in (7) becomes

$$\mathbf{w}_i - \mathbf{w}_{i-1} = \mu\phi_i(\mathbf{x}_i - y_i\mathbf{w}_{i-1}) + O(\mu^2) \tag{9}$$

Taking expectation in (9) and using $E[\mathbf{w}_i] = E[\mathbf{w}_{i-1}] = \mathbf{w}$ in the steady-state, we obtain

$$\mathbf{0} = \mu E[\phi_i(\mathbf{x}_i - y_i\mathbf{w})] + E[O(\mu^2)] \tag{10}$$

Ignoring $E[O(\mu^2)]$ for a small $\mu$, the steady-state condition in (10) approximates to

$$E[\phi_i(\mathbf{x}_i - y_i\mathbf{w})] \approx \mathbf{0} \tag{11}$$

Multiplying $\mathbf{w}$ in both sides of (11), we get

$$E[\phi_i(\mathbf{w}^T\mathbf{x}_i - y_i\mathbf{w}^T\mathbf{w})] = E[\phi_i(y_i - y_i\|\mathbf{w}\|^2)] \approx 0 \tag{12}$$

TABLE I
COMPARISON OF COMPUTATIONAL COMPLEXITY

| Algorithms | $\times$ | $\div$ | $\sqrt{\cdot}$ |
|---|---|---|---|
| unit-norm in (3)-(4) | $K$ | $K$ | 1 |
| proposed in (5) | 1 | $K$ | 0 |

Also the equation in (12) can be rewritten as

$$E[\phi_i y_i](1 - \|\mathbf{w}\|^2) \approx 0 \tag{13}$$

From (13), in the steady-state, we can easily check that the proposed algorithm satisfies the unit-norm constraint

$$\|\mathbf{w}\|^2 = 1$$

or

$$E[\phi_i y_i] = 0.$$

In addition to the unit-norm constrain, the propose algorithm has the other stationary point, $E[\phi_i y_i] = 0$. Although we do not provide the rigorous analysis for the condition $E[\phi_i y_i] = 0$, the extensive simulation results indicate that the algorithm hardly results in the undesired solution.

Table I describes the computational complexity required for performing the normalization steps.

## III. Applications

To demonstrate the proposed method indeed solves the constrained optimization problem, we apply the method to two well-known signal processing problems: constrained mean-square optimization and principle component analysis (PCA).

### A. Constrained Mean-Square Optimization

Here we consider

$$J(\mathbf{w}) = f(e_i) = E[e_i^2] \tag{14}$$

Then

$$\frac{\partial J(\mathbf{w})}{\partial \mathbf{w}_{i-1}} = -E[e_i \mathbf{x}_i] \tag{15}$$

Taking instantaneous value of $E[e_i \mathbf{x}_i]$ and comparing (15) with (3),

$$\phi_i = e_i \tag{16}$$

Therefore, the proposed adaptive algorithm in (5) becomes

$$\mathbf{w}_i = \frac{\mathbf{w}_{i-1} + \mu e_i \mathbf{x}_i}{1 + \mu e_i y_i}. \tag{17}$$

The desired data $d_i$ is modelled as the output of the finite impulse response (FIR) filter, $\mathbf{w}^\circ$

$$d_i = \mathbf{w}^{\circ T} \mathbf{x}_i + v_i \tag{18}$$

where we set $\mathbf{w}^\circ = [1 \quad 0.5 \quad 0.2 \quad -0.5 \quad 0.3]^T$ and the zero mean white Gaussian noise is added such that the signal-to-noise ratio is $20dB$. We want to find $\mathbf{w}$ such that

$$\min_{\mathbf{w}} (J(\mathbf{w}) = f(e_i) = E[e_i^2]) \tag{19}$$

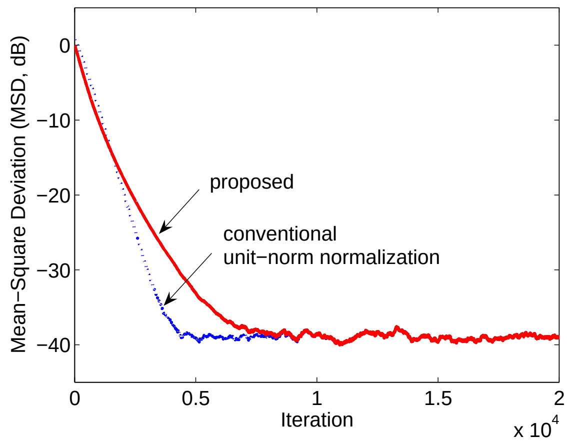


Fig. 1. Performance of two adaptive algorithms in constrained mean-square optimization (conventional vs. proposed)

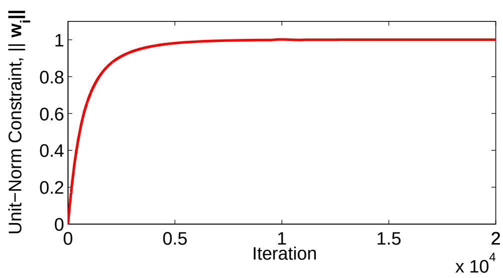


Fig. 2. Unit-norm constraint of the proposed algorithm, $\|\mathbf{w}_i\|$.

subject to

$$\|\mathbf{w}\|^2 = \sum_{k=0}^{K-1} |w(k)|^2 = 1.$$

The optimal solution of (19) is

$$\mathbf{w}^* = \mathbf{w}^\circ / \|\mathbf{w}^\circ\|. \tag{20}$$

In the adaptive algorithm in (17) we choose the adaptive filter length $K = 4$, and $\mu = 0.001$. The initial condition is set to $\mathbf{w}_0 = [0 \quad 0 \quad 0 \quad 0 \quad 0]^T$. For comparison purposes, the conventional algorithm in (3) and (4) is also tested using the same model.

Fig. 1 shows the mean-square deviation (MSD) of two algorithms,

$$\text{MSD} = 10 \log_{10} \|\mathbf{w}^* - \mathbf{w}_{i-1}\|^2.$$

Each plot is averaged over 100 independent trials. Dashed line is the result of the conventional algorithm and the solid is that of the proposed. As can be seen in Fig. 1, the proposed works as good as the conventional while keeping low computational complexity. We also plot $\|\mathbf{w}_i\|$ in Fig. 1 checking the unit-norm constraint.

### B. Principle Component Analysis (PCA)

PCA is a well-established technique for dimension reduction. Its applications include data compression, image processing, data visualization, pattern recognition, and time series prediction [2]–[4]. Let $\mathbf{x} \in \Re^K$ denote an $K$-dimensional zero mean random vector. Consider a single neural network whose output $y_i$ is given by

$$y_i = \mathbf{w}_{i-1}^T \mathbf{x}_i. \tag{21}$$

The PCA aims at finding a vector $\mathbf{w}_i$ such that the variance of $y_i$ is maximized, i.e.,

$$\mathbf{w} = \arg\max E[y_i^2] \tag{22}$$

subject to

$$\|\mathbf{w}\| = 1$$

. It is well-known that the solution $\mathbf{w}$ corresponds to the normalized eigenvector associated with the largest eigenvalue of the covariance matrix $R_{xx} = E[\mathbf{x}_i \mathbf{x}_i^T]$. A conventional adaptive algorithm finding $\mathbf{w}$ is

$$\begin{aligned} \tilde{\mathbf{w}}_i &= \mathbf{w}_{i-1} + \mu y_i \mathbf{x}_i, \quad \text{and} \\ \mathbf{w}_i &= \frac{\tilde{\mathbf{w}}_i}{\|\tilde{\mathbf{w}}_i\|}. \end{aligned} \tag{23}$$

Here we can easily derive a simple PCA algorithm based on single parameter normalization. From (22),

$$J(\mathbf{w}) = f(e_i) = E[y_i^2]$$

Then

$$\frac{\partial J(\mathbf{w})}{\partial \mathbf{w}_{i-1}} = y_i \mathbf{x}_i \tag{24}$$

Comparing (24) with (3), we get

$$\phi_i = y_i \tag{25}$$

Therefore, replacing $\phi_i$ with $y_i$ in (5) the proposed adaptive algorithm for PCA becomes

$$\mathbf{w}_i = \frac{\mathbf{w}_{i-1} + \mu y_i \mathbf{x}_i}{1 + y_i^2}. \tag{26}$$

Using the approximation in (9), the algorithm in (26) reduces to

$$\mathbf{w}_i = \mathbf{w}_{i-1} + \mu(y_i \mathbf{x}_i - y_i^2 \mathbf{w}_{i-1}) + O(\mu^2) \tag{27}$$

Neglecting $O(\mu^2)$ for a small $\mu$, the algorithm in (27) is exactly the same as the well-known Oja's rule [8]–[9]. Also the steady-state condition in (13) becomes

$$E[\phi_i y_i](1 - \|\mathbf{w}\|^2) = E[y_i^2](1 - \|\mathbf{w}\|^2 \approx 0 \tag{28}$$

Since $E[y_i^2] = \mathbf{w}^T R_{xx} \mathbf{w} > 0$, the only solution of (28) is

$$\|\mathbf{w}\|^2 = 1.$$

In the simulation the input vector $\mathbf{x}$ is chosen to have the covariance matrix of

$$R_{xx} = \begin{pmatrix} 1.00 & 0.80 & 0.59 & 0.30 \\ 0.80 & 1.00 & 0.60 & 0.49 \\ 0.59 & 0.60 & 1.00 & 0.80 \\ 0.30 & 0.49 & 0.80 & 1.00 \end{pmatrix},$$

whose the normalized eigenvector associated with the largest eigenvalue is

$$\mathbf{w}^* = [0.4617 \;\; 0.5356 \;\; 0.5356 \;\; 0.4616]^T.$$

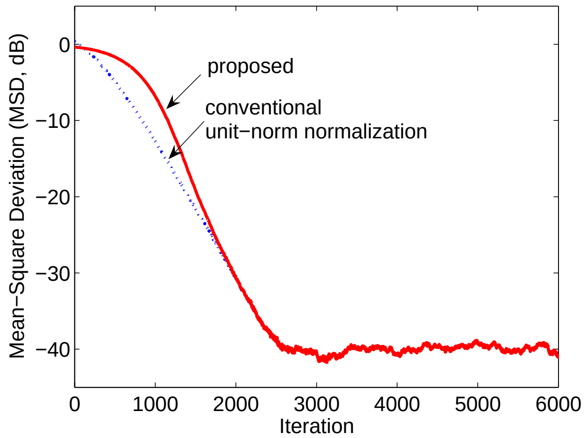


Fig. 3. Performance of two adaptive algorithms in PCA (conventional vs. proposed method)

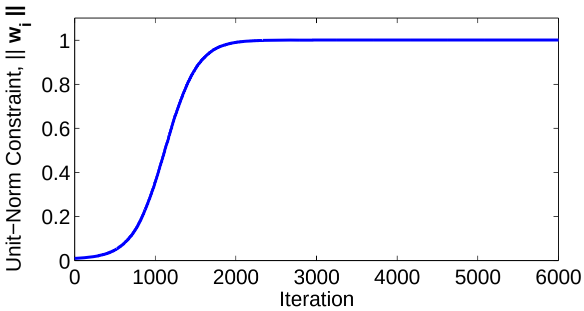


Fig. 4. Unit-norm constraint of the proposed PCA algorithm, $\|\mathbf{w}_i\|$.

Also the initial condition is set to $\mathbf{w}_0 = [0.1 \;\; 0 \;\; 0 \;\; 0 \;\; 0]^T$. We set $K = 2$ and $\mu = 0.001$. For comparison we plot the norm squared parameter error, $\|\mathbf{w}^* - \mathbf{w}_{i-1}\|^2$. Each plot is averaged over 100 independent trials.

As shown in Fig. 3, the proposed adaptive algorithm converges as good as the conventional one while being computationally simpler. We also plot $\|\mathbf{w}_i\|$ in Fig. 4 to check the unit-norm constraint.

## IV. Conclusions

We have proposed an simple adaptive algorithm for the unit-norm constrained optimization. From simulation results, the proposed scalar normalization works as good as the conventional norm-based normalization while being computationally simpler. Although only two applications have been illustrated here, we believe the proposed method may trigger other applications based on the unit-norm constrained optimization.

## References

[1] B. Widrow and S. D. Sterns, *Adaptive Signal Processing*, Englewood Cliffs, NJ: Prentice Hall, 1985.
[2] A. Benveniste, M. Metivier and P. Priouret, *Adapive Algorithms and Stochstic Approximation*, New York: Springer-Verlag, 1990.
[3] S. Haykin, *Adaptive Filter Theory*, Englewood Cliffs, NJ: Prentice Hall, 2002.

[4] A. H. Sayed, *Fundamentals of Adaptive Filtering*, Englewood Cliffs, NJ: Prentice Hall, 2003.
[5] S. C. Douglas, S. Amari and S. Y. Kung, "On gradient adaptation with uni-norm constraints," *IEEE Trans. Signal Processing*, vol. 48, no. 6, pp. 1843–1847, June 2000.
[6] S. G. Sankaran and A. A. Louis Beex, "Hyperspherical parametrization for unit-norm based adaptive IIR filtering," *IEEE Signal Processing Letters*, vol. 6, no. 12, pp. 318–320, Dec. 1999.
[7] R. López-Valcare, "An algorithm for unit-norm equation error system identification based on the method of multipliers," *IEEE Tran. on Signal Processing Letters*, vol. 6, no. 12, pp. 3080–3085, Dec. 2003.
[8] P. Comon, "Independent component analysis: A new concept?," *Signal Processing*, vol. 36, no. 3, pp. 287–314, Apr. 1994.
[9] A. Hyvarinen and E. Oja, "Independent component analysis by general nonlinear Hebbian-like learning rules," *Signal Processing*, vol. 64, no. 3, pp. 301–313, Feb. 1998.