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 \( \mathbf{w} = [w(0) \ w(1) \ \cdots \ w(K-1)]^T \) is to minimize a quadratic cost function of an error

\[
e_i = (d_i - \mathbf{w}^T \mathbf{x}_i)^2
\]

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

\[
\min_{\mathbf{w}} J(\mathbf{w}) = f(e_i)
\]

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

\[
\hat{\mathbf{w}}_i = \mathbf{w}_{i-1} - \mu \frac{\partial f(e_i)}{\partial \mathbf{w}_{i-1}}
\]

\[
\mathbf{w}_i = \frac{\hat{\mathbf{w}}_i}{\|\hat{\mathbf{w}}_i\|}
\]

where \( \mathbf{w}_i \) is an estimate at iteration \( i \), \( \mu \) is the step-size, \( \phi_i = \frac{\partial f(e_i)}{\partial \mathbf{w}_{i-1}} \) 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 \( \mathbf{w}_i \) so that the algorithm satisfies \( \|\mathbf{w}_i\| = 1 \) at every iteration.

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:

\[
\hat{\mathbf{w}}_i = \mathbf{w}_{i-1} + \mu \phi_i \mathbf{x}_i, \quad \text{and} \quad \mathbf{w}_i = \frac{\mathbf{w}_i}{1 + \mu \phi_i y_i},
\]

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

\[
\mathbf{w}_i = \mathbf{w}_{i-1} + \frac{\mu \phi_i \mathbf{x}_i}{1 + \mu \phi_i y_i}.
\]

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

\[
\mathbf{w}_i - \mathbf{w}_{i-1} = \frac{\mathbf{w}_{i-1} + \mu \phi_i \mathbf{x}_i - \mathbf{w}_{i-1}}{1 + \mu \phi_i y_i} = \frac{\mu \phi_i \mathbf{x}_i - y_i \mathbf{w}_{i-1}}{1 + \mu \phi_i y_i}.
\]

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

\[
\frac{1}{1 + \mu \phi_i y_i} = 1 + O(\mu),
\]

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

\[
0 = \mu E[\phi_i (x_i - y_i w)] + E[O(\mu^2)]. \tag{10}
\]

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

\[
E[\phi_i (x_i - y_i w)] \approx 0. \tag{11}
\]

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

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

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 (shnhc@ssu.ac.kr).
Also the equation in (12) can be rewritten as
\[ E[\phi_i y_i](1 - \|\mathbf{w}\|^2) \approx 0 \] (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
\[ J(\mathbf{w}) = f(\mathbf{w}) = E[e_i^2] \] (14)
Then
\[ \frac{\partial J(\mathbf{w})}{\partial \mathbf{w}_{i-1}} = -E[e_i x_i] \] (15)
Taking instantaneous value of \( E[e_i x_i] \) and comparing (15) with (3),
\[ \phi_i = e_i \] (16)
Therefore, the proposed adaptive algorithm in (5) becomes
\[ \mathbf{w}_i = \frac{\mathbf{w}_{i-1} + \mu e_i \mathbf{x}_i}{1 + \mu e_i y_i}. \] (17)
The desired data \( d_i \) is modelled as the output of the finite impulse response (FIR) filter, \( \mathbf{w}_0 \)
\[ d_i = \mathbf{w}_0^T \mathbf{x}_i + v_i \] (18)
where we set \( \mathbf{w}_0 = \begin{bmatrix} 1 & 0.5 & 0.2 & -0.5 & 0.3 \end{bmatrix}^T \) 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
\[ \min_{\mathbf{w}}(J(\mathbf{w}) = f(\mathbf{w}) = E[e_i^2]) \] (19)

#### 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 \( x \in \mathbb{R}^K \) denote an \( K \)-dimensional zero mean random vector. Consider a single neural network whose output \( y_i \) is given by
\[
y_i = w_{i-1}^T x_i.
\]
(21)
The PCA aims at finding a vector \( w_i \) such that the variance of \( y_i \) is maximized, i.e.,
\[
w = \arg \max E[y_i^2]
\]
subject to
\[\|w\| = 1\] . It is well-known that the solution \( w \) corresponds to the normalized eigenvector associated with the largest eigenvalue of the covariance matrix \( R_{xx} = E[x_i x_i^T] \). A conventional adaptive algorithm finding \( w \) is
\[
w_i = w_{i-1} + \mu y_i x_i, \quad \text{and} \quad w_i = \frac{w_i}{\|w_i\|}.
\]
(23)
Here we can easily derive a simple PCA algorithm based on single parameter normalization. From (22),
\[
J(w) = f(\epsilon_i) = E[y_i^2]
\]
Then
\[
\frac{\partial J(w)}{\partial w_{i-1}} = y_i x_i
\]
(24)
Comparing (24) with (3), we get
\[
\phi_i = y_i
\]
(25)
Therefore, replacing \( \phi_i \) with \( y_i \) in (5) the proposed adaptive algorithm for PCA becomes
\[
w_i = \frac{w_{i-1} + \mu y_i x_i}{1 + y_i^2}.
\]
(26)
Using the approximation in (9), the algorithm in (26) reduces to
\[
w_i = w_{i-1} + \mu (y_i x_i - y_i^2 w_{i-1}) + O(\mu^2)
\]
(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
\[
E[\phi_i y_i](1 - \|w\|^2) = E[y_i^2](1 - \|w\|^2) \approx 0
\]
(28)
Since \( E[y_i^2] = w^T R_{xx} w > 0 \), the only solution of (28) is
\[\|w\|^2 = 1\]. In the simulation the input vector \( x \) 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
\[
w^* = [0.4617 \ 0.5356 \ 0.5356 \ 0.4616]^T.
\]


