CONVERGENCE BEHAVIOUR OF LMS-TYPE ALGORITHMS FOR ADAPTIVE NOISE CONTROL IN NOISY DOPPLER ENVIRONMENTS

ROBERT W. STEWART, STEPHAN WEISS, DAVID H. CRAWFORD

Signal Processing Division, Dept. of Electronic and Electrical Engineering, University of Strathclyde, Glasgow G1 1XW, Scotland, UK

Abstract. This paper discusses the convergence and tracking behaviour of LMS-type algorithms in a certain type of environment, which is characterised by a Doppler shift in frequency between the two signals available to the algorithm and rapid variations in signal power. We show the linear time-varying characteristics of the underlying system and derive optimum trajectories to which we can compare the adaptation and tracking ability of first order LMS and NLMS adaptive filters. We also present simulations using higher filter orders and real world noise, for which particular emphasis is put on the presence of observation noise. An excursion into the theory of non-stationary convergence and tracking of adaptive algorithms provides justification for the observed behaviour of the algorithms.

Key Words. Non-stationary adaptive filtering, adaptive noise cancellation, active noise cancellation.

1. INTRODUCTION

As for most adaptive filtering applications, adaptive and active noise cancellation (ANC) can be reduced to an identification problem. The solid part of Fig. 1 represents the essential blocks of a single channel active noise cancellation architecture, where the adaptive filter is adjusted in a way that its output signal $y(k)$, fed into a physical system, interferes with a desired signal $d(k)$ such that the resulting error is minimum [4]. Adaptive noise cancellation possesses a similar structure, with the auxilary path h being minus unity and the interference taking place as an arithmetic subtraction $e(k) = d(k) - y(k)$ [21]. Both applications try to suppress an unwanted noise $n'(k)$ by appropriately filtering a "similar" input or reference signal $x(k) = n(k)$. The origin of this similarity is marked in Fig. 1 as a dashed part, where $n'(k)$ is related to $n(k)$ by an unknown system f. Ideally,

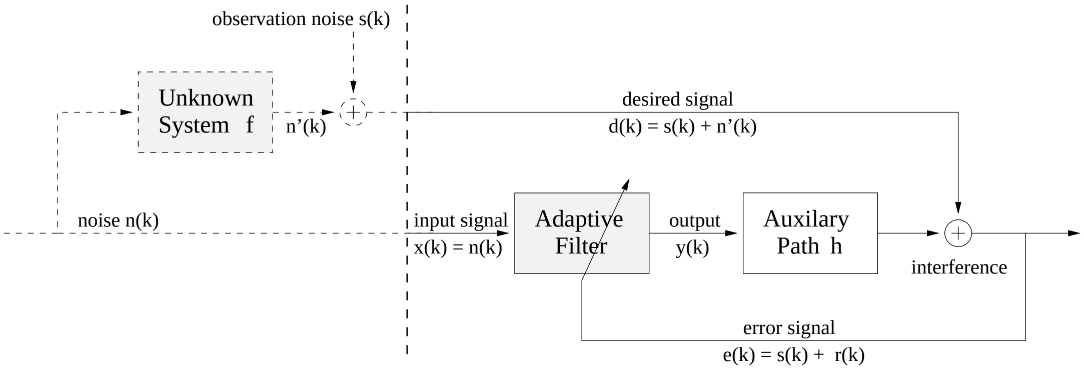

Fig. 1: generic structure of single channel adaptive or active noise cancellation

this system f has to be identified by the adaptive filter, which in the case of ANC additionally has to determine the inverse of the auxilary path h. Generally, f can be non-linear, time-variant, and partly non-causal. Therefore for the feasibility of adaptive and/or active noise cancellation, and the subsequent selection of the adaptive filter, some a priori knowledge about the characteristics of the unknown system f is important, as eg. linear adaptive filters can only identify the causal part of the cross-correlation between $n(k)$ and $n'(k)$.

We have recently come across an unclear situation of the unknown system f, where adaptive or active noise cancellation are to be applied to signals picked up by spatially separated microphones in an environment in which either the (primary) noise source or the filter microphones / ANC setup are moved past each other, thus creating different Doppler shifts in frequency and different variations in the signal power in the signals being supplied to the adaptive system. A number of application examples exhibiting such characteristics include

- ANC at the side of noisy vehicle movement, eg. roads, railway tracks, runways etc.,
- ANC inside vehicles moving past a noise source being either stationary or moving at different speed,
- adaptive noise cancellation for telephones along roads or railway tracks, and
- adaptive noise cancellation for communication line enhancement as eg. mobile telephones in cars etc. moving past a noise source.

A mathematical analysis of such scenarios has been performed in [19], revealing the linear, time-varying nature of the underlying unknown system f. Therefore, linear adaptive filters are suitable to apply if they are able to track non-stationarities fast enough.

For adaptive filters using the least-mean-square (LMS) algorithm, the tracking ability in non-stationary environments has been shown to be governed by a trade-off between lagging behind the optimum solution and the amount of gradient noise introduced by the step size [20, 10, 8]. Noise-free simulations in [19] agree well with theory and fast LMS versions like the normalised LMS (NLMS)[15] show greatly enhanced tracking and cancellation results.

However, all the above mentioned application examples include a potential high level of observation noise $s(k)$ – mostly speech, as eg. passenger communication inside cars – or a useful audio signal, which has to be enhanced by reducing present noise $n'(k)$ to a residual noise $r(k)$ in the error signal. This observation noise will distract adaptation. At the same time it must not be affected by the adaptive system in order to avoid distortions.

In the following, we will give some insight into the mathematical model of Doppler shifted noise signals, the derivation of the underlying unknown system f, and the optimum trajectories of the coefficients of a discrete-time first order filter. For applying adaptive filtering using the least-mean-square (LMS) and normalised LMS algorithms, some of the important characteristics from theory are presented in Sec. 3, which will help to explain the convergence and tracking behaviour observed in Sec. 4, with particular reference to observation noise. As the important part of this work refers to the identification of a certain type of non-stationarity, we will restrict the analysis to adaptive noise cancellation, but will briefly discuss the additional problems involved for ANC in Sec. 5.

2. PROBLEM DESCRIPTION AND ANALYSIS

2.1. Received Signals

For the analysis, we assume a single point noise source, S, moving with constant velocity, $\underline{v} = \hat{v}\underline{e}_x$, emitting a sinusoidal signal $p(t) = \hat{p}\sin(\omega t)$ of constant frequency ω and amplitude $\hat{p}$, producing the vectorised model of the environment shown in Fig. 2. The noise signal received at two stationary microphone positions M_1 and M_2 has the form

$$p_i(t) = \frac{\hat{p}}{r_i(t)} \cdot \sin\left(\omega t - k r_i(t)\right), \quad i \in \{1,2\}, \qquad (1)$$

where r_i are the distances the sound travels from the instance of emission until reception, and $k = \omega/c$ is the wavenumber and c the velocity of sound in air. The time-variant phase causes a Doppler shift in frequency and therefore a difference in instantaneous frequency between both signals, the extent of which depends on the separation of the reception points. Also note that the received signal power depends on $1/r_i^2$ due to the attenuation in air. For the calculation, we introduce the time duration

$$\Theta_i = \frac{r_i}{c}, \quad i \in \{1,2\}, \qquad (2)$$

which the signal received at time t needed to travel since it had been emitted.

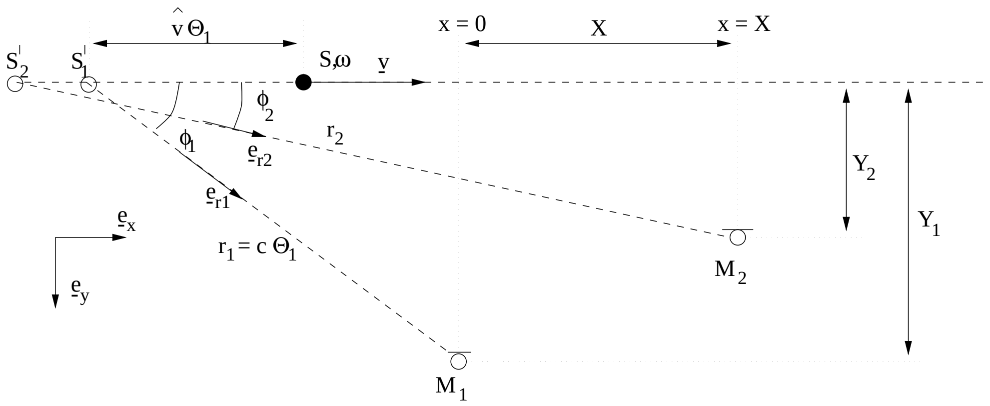

Fig. 2: general model of a Doppler shift producing environment

Using geometric considerations from Fig. 2, we find

$$\begin{aligned} r_1 &= \Theta_1 \cdot c = \sqrt{\left(x - \hat{v}\Theta_1\right)^2 + Y_1^2} \quad , \quad \text{and} \\ r_2 &= \Theta_2 \cdot c = \sqrt{\left(x - X - \hat{v}\Theta_2\right)^2 + Y_2^2} \quad , \end{aligned}$$

where X is the horizontal separation between the two microphones and Y_1 and Y_2 their distances from the track, and $x(t)$ the current position of the vehicle. Solving the quadratic equations for $\Theta_i, i \in 1, 2$ and using the restriction imposed by a causal system of $\Theta_i \geq 0$ yields

$$\begin{aligned} \Theta_1 &= \frac{x\hat{v} + \sqrt{x^2\hat{v}^2 + \left(x^2 + Y_1^2\right)\left(c^2 - \hat{v}^2\right)}}{c^2 - \hat{v}^2}, \quad \text{and} \\ \Theta_2 &= \frac{-(x–X)\hat{v}}{c^2 - \hat{v}^2} + \\ &\quad + \frac{\sqrt{(x–X)^2\hat{v}^2 + \left((x–X)^2 + Y_2^2\right)\left(c^2 - \hat{v}^2\right)}}{c^2 - \hat{v}^2} \end{aligned}$$

with $x = \hat{v}t$ being time dependent. Exploiting (2) now yields all the parameters for the received signals $p_i(t)$ of (1) in terms of the geometric arrangement $\{X, Y_1, Y_2\}$ and the vehicle parameters $\{\hat{v}, \omega\}$.

2.2. Consistency with Doppler

Differentiating the arguments of the sine terms in (1) with respect to time t

$$\begin{aligned} \frac{d}{dt}\left(\omega t - \frac{\omega}{c} r_i\right) &= w\left(1 - \frac{1}{c} \cdot \frac{dr_i}{dt}\right) \\ &= \omega \cdot \left(1 - \frac{v_i}{c}\right) = \tilde{\omega}_i, \quad i \in \{1, 2\} \quad (3) \end{aligned}$$

we yield a term $\tilde{\omega}_i$ commonly known as Doppler frequency, where v_1 and v_2 are the velocity components of the source velocity $\underline{v}$ at the virtual locations S'_1 and S'_2, where the signal received at time t had been emitted, in direction of the microphones:

$$\begin{aligned} v_1 &= \hat{v} \cos \phi_1 = \qquad (4) \\ &= \hat{v} \cos\left(\tan^{-1}\left(\frac{Y_1}{x - \hat{v}\Theta_1}\right)\right) \operatorname{sgn}(x - \hat{v}\Theta_1), \end{aligned}$$

$$v_2 = \hat{v} \cos\left(\tan^{-1}\left(\frac{Y_2}{x - X - \hat{v}\Theta_2}\right)\right) \operatorname{sgn}(x - X - \hat{v}\Theta_2).$$

Here, the signum function has been introduced to correct for the side branches of the arctangent. As the v_i are depending on time t, the terms

$$\tilde{\omega}_i = \omega\left(1 - \frac{v_i}{c}\right), \quad i \in \{1, 2\} \qquad (5)$$

have to be interpreted as instantaneous frequencies. Due to the different location of the microphones M_1 and M_2, these instantaneous frequencies are different for both recorded signals. For applying an adaptive filter, reference and desired or error signal would thus show a shift in instantaneous frequency due to the Doppler phenomenon of the received signals.

2.3. Derivation and Classification of the Underlying System

If adaptive filtering is employed to suppress p_2, a filter with input p_1 will ideally have to identify the function $f : p_2 = f(p_1)$. For the identification of f, a reformulation of p_2 in terms of p_1 can be performed:

$$\begin{aligned} p_2(t) &= \frac{\hat{p}}{r_2} \sin(\omega t - k r_2) \qquad (6) \\ &= \frac{\hat{p}}{r_2} \sin(\omega t - k r_1 - k(r_2 - r_1)) \\ &= \frac{r_1}{r_2} \cos(k(r_1 - r_2)) \cdot p_1(t) \quad + \\ &\quad + \frac{\hat{p}}{r_2} \cos(\omega t - k r_1) \cdot \sin(k(r_1 - r_2)) \quad (7) \end{aligned}$$

To relate the second summand directly to $p_1(t)$, we differentiate p_1 with respect to time t

$$\dot{p}_1(t) = -\frac{v_1}{r_1} \cdot p_1(t) + \hat{p}\frac{\omega(1 - v_1/c)}{r_1} \cos(\omega t - k r_1)$$

where the identity $\dot{r}_1 = v_1$ has been used, and hence

$$\cos(\omega t - k r_1) = \left(\dot{p}_1(t) + \frac{v_1}{r_1} \cdot p_1(t)\right) \cdot$$

$$\cdot\frac{1}{\hat{p}} \cdot \frac{r_1}{\omega(1-v_1/c)}. \qquad (8)$$

Inserting (8) into (7) yields as an expression for f a *linear* first order differential equation

$$p_2(t) = f(p_1) \;=\; a_0(t) \cdot \dot{p}_1(t) \;+\; a_1(t) \cdot p_1(t) \qquad (9)$$

with *time variant* parameters

$$\begin{aligned} a_0(t) &= \frac{r_1}{r_2}\frac{1}{\omega(1-v_1/c)}\sin(k(r_1-r_2)) && (10)\\ a_1(t) &= \frac{r_1}{r_2}\cos(k(r_1-r_2)) \;+\; \frac{v_1}{r_1} \cdot a_0(t) \quad . && (11) \end{aligned}$$

As (10) and (11) no longer includes ωt, there cannot be any further p_1 terms extracted from a_0 or a_1. Thus the functional context of the system (1) is completely governed by (9)-(11), revealing the *linear, time variant* nature of the function $f : p2 = f(p_1)$.

Based on the previous analysis, it is now possible to perform the filtering task with *linear* adaptive filters. The main question remaining is whether the adaptive algorithm can track the time variant parameters of the system.

2.4. Discrete Time Model and Filtering

If the sound pressure signals are acquired in discrete time, i.e. $t \to k \cdot T_s$, the resulting discrete time sequences are determined by a set of parameters $I\!P = \{X, Y_1, Y_2, \hat{v}, \omega, f_s\}$ consisting of

- local arrangement (X,Y_1, and Y_2),
- relative speed $\hat{v}$ of source,
- source angular frequency $\omega = 2\pi f$, and
- sampling frequency $f_s = 1/T_s$,

which yield discrete sequences $p_1[k]$ and $p_2[k]$. An adaptive filter applied to noise cancellation as discussed in Fig. 1 is supplied with these sampled pressure signals as reference and desired signal, such that $x[k] = n[k] = p_1[k]$ and $d[k] = n'[k] = p_2[k]$ for the noise free case $s[k] \equiv 0$.

2.5. Optimal Trajectory of a Discrete-Time First-Order Filter

For a first order filter, there is a unique set of optimum coefficients $w_{i,\mathrm{opt}}[k], i \in \{0, 1\}$, such that

$$y[k] \overset{!}{=} p_2[k] = w_{0,\mathrm{opt}} \cdot p_1[k] + w_{1,\mathrm{opt}} \cdot p_1[k\text{–}1] \qquad (12)$$

is satisfied, where the optimum coefficients can be evaluated in an approach analogous to Sec. 2.3 as

$$w_{0,\mathrm{opt}}[k] = \frac{r_1[k]\sin(k(r_2[k]-r_1[k\text{–}1])-\omega/f_s)}{r_2[k]\sin(k(r_1[k]-r_1[k\text{–}1])-\omega/f_s)} \qquad (13)$$

$$w_{1,\mathrm{opt}}[k] = \frac{-r_1[k\text{–}1]\sin(k(r_1[k]-r_2[k]))}{r_2[k]\sin(k(r_1[k]-r_1[k\text{–}1])-\omega/f_s)} \qquad (14)$$

Thus, the optimum filter has a dynamic, non-stationary solution. The shape of the optimum trajectories, $w_{0,\mathrm{opt}}$ and $w_{1,\mathrm{opt}}$, depends on the parameter set $I\!P$. An example for the curve of these trajectories is illustrated in Fig. 5(a,b). The strongest variations in the trajectories occur during the transition, when the noise source passes the microphones in $t = 0$, while for approach and departure stage the trajectories remain almost constant. Generally , the further apart the reception points M_1 and M_2 are, and the higher the relative speed $\hat{v}$ is, the stronger the non-stationarities become [19].

3. ADAPTIVE FILTERING IN NON-STATIONARY ENVIRONMENTS

3.1. Notations

Coming back on adaptive filtering as shown in Fig. 1, we define the delayed input values and the filter weights in vector notation

$$\begin{aligned} \mathbf{x}_k &= [x(k), x(k-1), \ldots x(k-N+1)]^T && (15)\\ \mathbf{w}_k &= [w_0(k), w_1(k), \ldots w_{N-1}(k)]^T && (16) \end{aligned}$$

to write the filter equation of an N-tap filter as

$$y(k) = \mathbf{x}_k^T \mathbf{w}_k \qquad (17)$$

where $(\cdot)^T$ denotes transpose. Considering adaptive noise cancellation, the error signal is then given by $e(k) = d(k) - y(k)$. For the following, $x(k)$ and $d(k)$ are assumed to be non-stationary, zero-mean processes.

3.2. Optimum Wiener Solution

An optimum solution for the filter weights $\mathbf{w}_k$ can be calculated using (13) and (14) for deterministic signals.A general approach, which has to be interpreted very carefully, is given by the Wiener solution [21, 8]

$$\mathbf{w}_{\mathrm{opt},k} = \mathbf{R}_k^{-1}\mathbf{p}_k \qquad (18)$$

where $\mathbf{R}_k$ and $\mathbf{p}_k$ are covariance matrix and cross-correlation vector. Special attention has to be drawn to the fact, that these are time dependent ensemble statistics, not time averages:

$$\begin{aligned} \mathbf{R}_k &= \mathcal{E}\{\mathbf{x}_k \cdot \mathbf{x}_k^T\} && (19)\\ \mathbf{p}_k &= \mathcal{E}\{\mathbf{x}_k \cdot d(k)\} && (20) \end{aligned}$$

with $\mathcal{E}\{\cdot\}$ denoting expectations. As for practical applications, $\mathbf{R}_k$ and $\mathbf{p}_k$ usually have to be estimated from time-averages, applying Wiener-Hopf should be restricted to strictly stationary problems [14].

3.3. Least-Mean-Square (LMS) Algorithm

The least-mean-square algorithm [21, 20] is an iterative, stochastic gradient descent based method updating the filter weights according to

$$\mathbf{w}_{k+1} = \mathbf{w}_k + 2\mu\mathbf{x}_k e(k) \quad . \tag{21}$$

Convergence in the mean, ie. $\mathbf{w}_k \rightarrow \mathbf{w}_{\mathrm{opt}}$ for $k \rightarrow \infty$, is analysed in detail eg. in [20, 8]. Using some assumptions over the weight changes and a translation and rotation of the coordinate system

$$\mathbf{v}'_k = \mathbf{Q}_k^T \left(\mathbf{w}_k - \mathbf{w}_{\mathrm{opt},k}\right) \quad , \tag{22}$$

where the matrix $\mathbf{Q}_k$ stems from the modal decomposition of the covariance matrix

$$\begin{aligned} \mathbf{R}_k &= \mathbf{Q}_k \mathbf{\Lambda}_k \mathbf{Q}_k^T \quad \text{, with} \\ \mathbf{Q}_k &= \text{modal matrix of } \mathbf{R}_k \quad \text{, and} \\ \mathbf{\Lambda}_k &= \mathrm{diag}\{\lambda_{i,k}\} \ , \ \lambda_{i,k} \text{ eigenvalues of } \mathbf{R}_k \ , \end{aligned}$$

the LMS update (21) can be decoupled to

$$\mathbf{v}'_{k+1} = \left(\mathbf{I} - 2\mu\mathbf{\Lambda}_k\right)\mathbf{v}'_k \quad . \tag{23}$$

Thus, a stability bound for the step size parameter μ governing convergence speed and final misadjustment can be derived

$$0 < \mu < \frac{1}{N \cdot \sigma^2_{xx,k}} \leq \frac{1}{\lambda_{\mathrm{max},k}} \tag{24}$$

depending on the largest eigenvalue of $\mathbf{R}_k$, which can be estimated by the variance of the input signal, $\sigma^2_{xx,k}$. Note again that due to the non-stationarity of $x(k)$, $\sigma^2_{xx,k}$ refers to the ensemble statistics and is not a time average.

For ANC, the LMS has to be modified to a filtered-X version [21, 4], which shows slower convergence and an increased error power [3, 18].

3.4. Normalised LMS Algorithm

As for many application $\sigma^2_{xx,k}$ cannot be obtained a priori, a normalisation of the LMS is introduced

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \frac{2\tilde{\mu}\mathbf{x}_k e(k)}{\mathbf{x}_k^T\mathbf{x}_k} \quad , \tag{25}$$

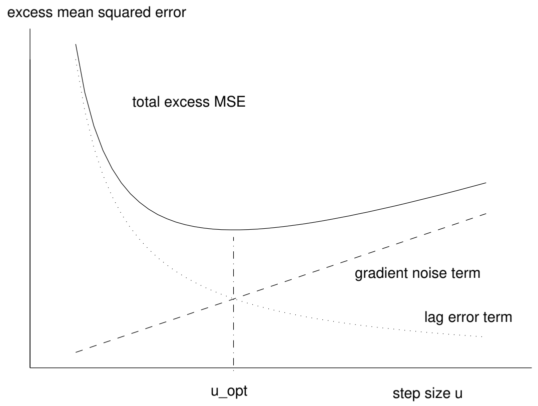

Fig. 3: misadjustment of the LMS algorithm in a non-stationary environment in dependency of the step size parameter μ.

which adjusts the step size by an estimate $\mathbf{x}_k^T\mathbf{x}_k$ of the input signal power [7, 1, 15]. For ANC applications, the filtered-X LMS can also be normalised to improve convergence speed [3, 13].

3.5. Convergence and Tracking Behaviour

The tracking behaviour of the LMS can be viewed in terms of the mean-square deviation of the filter coefficients from the optimum and the excess means square error (MSE) [20, 10, 8]. Based on the deviation of the weights

$$\begin{aligned} \mathbf{v}_k &= \mathbf{w}_k - \mathbf{w}_{\mathrm{opt},k} = \\ &= \big(\underbrace{\mathbf{w}_k - \mathcal{E}\{\mathbf{w}_k\}}_{\text{gradient noise}}\big) + \big(\underbrace{\mathcal{E}\{\mathbf{w}_k\} - \mathbf{w}_{\mathrm{opt},k}}_{\text{lag error}}\big), \end{aligned}$$

which can be separated into two different error terms, both mean-square deviation $\mathcal{D}_k = \mathcal{E}\{||\mathbf{v}_k||_2\}$ and excess MSE $\xi_{\mathrm{ex},k} = \mathcal{E}\{e^2(k)\} - \xi_{\mathrm{min}}$, where ξ_{min} is the min. MSE, can be decoupled [8, 10]

$$\mathcal{D}_k \simeq \mu N \sigma^2_{ss,k} + \frac{1}{4\mu}\mathrm{tr}\{\mathbf{R}_k^{-1}\mathbf{O}_k\} \tag{26}$$

$$\xi_{\mathrm{ex},k} \simeq \mu\sigma^2_{ss,k}\mathrm{tr}\{\mathbf{R}_k\} + \frac{1}{4\mu}\mathrm{tr}\{\mathbf{O}_k\} \quad , \tag{27}$$

with $\mathrm{tr}\{\cdot\}$ denoting trace of a matrix and $\mathbf{O}_k$ being the correlation matrix of the optimum weight changes. Thus, the first gradient noise term is proportional to μ, while the second term, which can be related to the lag error, is inversely proportional to the step size. The resulting trade-off for the step size between gradient noise and lag error terms is shown in Fig. 3 for the excess MSE. For certain type of non-stationarities, the optimum step size μ_{opt} can be determined [2]. Also note, that observation noise almost only influences the gradient noise, not the lag error term [16].

In literature, there usually is a clear separation between convergence and tracking ability of an algorithm [8]. However, most of the analysed problems assume system changes which are stationary

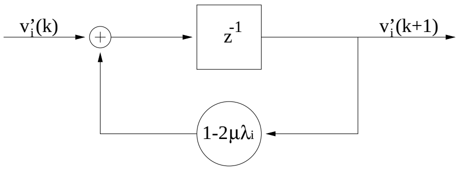

Fig. 4: integrator model for an LMS transformed weight update.

themselves [10, 20] and Eqns. (26) and (27) are steady-state quanitities [5]. However, the previously described Doppler environment ranges from stationary to strongly time variant behaviour, as can be seen from Fig. 6(a,b), and inhibits such a clear separation.

3.6. Estimation Approach to Convergence and Tracking

Intuitively, the LMS update (21) can be seen as a lowpass filter operation on the weights [8]. This can easily be verified, as the transformed update (23) forms a first order integrator on the decoupled weights, as shown in Fig. 4, with a time constant

$$\tau_{i,k} = \frac{1}{2\mu\lambda_{i,k}} \qquad (28)$$

also know as convergence speed in the mean [20], for each of the $i = 1 \ldots N$ weights. The time constant τ can be interpreted as the "memory" or data window of the weight update procedure. As for the LMS the step size μ will have to be adjusted to a rather small value due to large variations in the variance $\sigma^2_{xx,k}$, a longer time constant is yielded, while a shorter time constant can be assigned to the NLMS which always adjusts μ accordingly.

The task of the adaptive filter weights is to estimate the coefficients of the cross-correlation function between input and desired signal. Obviously, this statistical estimate has to be calculated via time averages, which even for stationary, ergodic processes requires a certain measurement time to yield confident estimates [14]. The arising dilemma is characterised by two extremes: while long measurement times due to small μ yield only estimates of restricted validity due to the non-stationarities, short measurement times due to a large step size will fail to reach any confidence.

For a small step size, the estimation process is expected to perform a lowpass operation on the true variations of the cross-correlation, thus averaging out fast parameter changes and distractions from the observation noise in the system. In contrary, for a large step size and therefore short measurement time, the estimation process cannot yield a