Convergence and the O.D.E. Method

David Watts

The University of Melbourne

9 November 2020

Table of Contents

Preliminaries

Applications I

The ODE Method

Applications II

Preliminaries

A general adaptive form

The context for this presentation is the following equation

$$\theta_n = \theta_{n-1} + \gamma_n H\left(\theta_{n-1}, X_n\right) + \gamma_n^2 \varepsilon_n\left(\theta_{n-1}, X_n\right) \tag{1}$$

This is a general form of a discrete-time recursive forumlation of the identification problem. A number of popular algorithms used in system identification can be obtained as special cases.

Symbols and their meaning

- $(\theta_n)_{n\geq 0}$ is the sequence of vectors to be recursively updated;
- $(X_n)_{n\geq 1}$ is a sequence of random vectors representing the on-line observations of the system in the form of a state vector;
- $(\gamma_n)_{n\geq 1}$ is a sequence of "small" scalar gains;
- $H(\theta, \bar{X})$ is the function which defines how the parameter θ is updated as a function of new observations;
- $\varepsilon_n(\theta, X)$ is a small perturbation term[1]

[1] $\varepsilon_n \equiv 0$ for algorithms discussed in this presentation

Analysis

Analysis of the equation is made difficult by the correction term. $\theta_n - \theta_{n-1}$ is implicitly dependent on past values of the parameters. An alternative route to analysis is simulation but this does not necessarily have generally applicability.

We are referring to an analysis of asymptotic convergence.

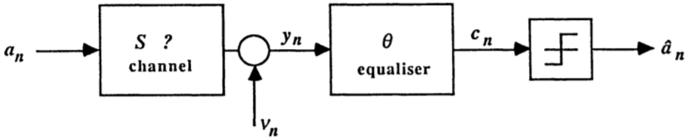
An analysis idea

It turns out that convergence properties will depend on approximations used for the gradient of the prediction. This is suggestive of the following idea: Take a discrete-time recursive algorithm and associate it with an ordinary differential equation. Then describe the asymptotic convergence of the algorithm in terms of the limiting behaviour of the ODE.

Applications I

Problem Description

For purposes of designing a recursive identification algorithm consider the following problem:



- Message a_n transmitted over unknown channel S
- Estimate $\hat{a}_n$ by filtering noisy observations y_n
- How to design θ with a low error rate?

Problem Description

If the inverse exists and there is zero measurement error then the ideal filter θ^* should be easy to find. There are two distinct phases[2] to the problem

1. learning: approximate θ using a known training sequence a_n
2. tracking: also known as self-adaptive equalisation where $\hat{\theta}$ is continuously updated to track changes in the channel due to exogenous factors

[2] A third phase called blind equalisation is required in the context of broadcast systems but will not be considered in this discssion

Model Structure: θ

A filter is generally the preferred structure for modelling the dynamical system of the equalisation problem. The filter structure opted for this presentation is the transveral (all-zeros) form.

$$c_n = \sum_{k=-N}^{+N} \theta(k) y_{n-k} \quad = Y_n^T \theta \tag{2}$$

where

$$Y_n^T := (y_{n+N}, \ldots, y_n, \ldots, y_{n-N}) \tag{3}$$

$$\theta^T := (\theta(-N), \ldots, 0(0), \ldots, \theta(N)) \tag{4}$$

Model Structure: θ

The model should be chosen to best describe the input/output characteristics:

- Transversal filter is linear and homogeneous function of time
- Permits a simplified the learning procedure
- Suitable for learning a limited class of channels
- Objective is to minimise algortihm complexity for associating the ODE
- Self-adaptive equalisation demands a more complex model such as the pole-zero form[3]

[3] See Albert Benveniste, Michel Métivier and Pierre Priouret, *Adaptive algorithms and stochastic approximations*, vol. 22 (Springer Science & Business Media, 2012) page 18 for an example.

Signal Modelling: a_n and v_n

The properties of the input signals are also an important specification of the model.

- (v_n) is typically a sequence of i.i.d. random variables with zero mean and fixed variance.
- (a_n) is a sequence of independent random variables with a uniformly distributed symbol set that is stationary and has zero mean (for example $(\pm 1, \pm 3)$
- Uniform distribution of (a_n) offers better bandwidth utilisation

An Adaptive Algorithm

The validity of an instance of the model structure of (2) can be measured using the estimation error. The variance of the estimation error forms the criterion function to be minimised by the adaptive algortihm:

$$\min_{\theta} V(\theta) \tag{5}$$

where

$$V(\theta) = \frac{1}{2}\mathrm{E}\left[a_n - c_n\right]^2 \tag{6}$$

Since $V(\theta)$ is quadratic in θ a solution may be found from:

$$\left[-\frac{d}{d\theta}V(\theta)\right]^{\mathrm{T}} = \mathrm{E}Y_n\left[a_n - Y_n^T \cdot \theta_{n-1}\right] = 0 \tag{7}$$

An Adaptive Algorithm

This gives an optimal value for θ of

$$\theta_* = \left[E\left(Y_n Y_n^T\right)\right]^{-1} E\left(Y_n a_n\right) \tag{8}$$

Unfortunately the joint distribution of Y_n and a_n is not known a priori and therefore the expectation can not be evaluated.

An Adaptive Algorithm

Suppose instead we are given a sequence of random variables X_n, each drawn from the same distribution. Given a function $F(\theta, X_n)$ whose values can be observed or constructed for some value of θ, find a solution to

$$EF(\theta, X_n) = f(\theta) = 0 \tag{9}$$

An Adaptive Algorithm

With θ assuming the role of the parameter, and state vector defined as

$$X_n = \begin{pmatrix} a_n \\ Y_n \end{pmatrix} \tag{10}$$

a suitable choice for F is the gradient term

$$F(\theta, X_n) = Y_n \left[a_n - Y_n^T \cdot \theta_{n-1} \right] \tag{11}$$

which is easy to evaluate given observations of X_n and the most recent parameter estimate.

An Adaptive Algorithm

An approach to solving (9) is to choose a sequence X_n, observe values of $F(\theta, X_n)$ and to infer the solution. Changing the value of X_n on every iteration it turns out is more efficient. Robbins and Munro[4] suggested the following recursive scheme for solving (9):

$$\hat{\theta}_n = \hat{\theta}_{n-1} + \gamma_n F(\hat{\theta}_{n-1}, X_n) \tag{12}$$

[4]Herbert Robbins and Sutton Monro, 'A stochastic approximation method', *The annals of mathematical statistics* (1951): 400–407.

An Adaptive Algorithm

Application of this scheme to (7) gives

$$\hat{\theta}_n = \hat{\theta}_{n-1} + \gamma_n Y_n \left[a_n - Y_n^T \cdot \theta_{n-1} \right] \tag{13}$$

In adaptive signal processing this is widely known as the "Least Mean Squares" algortihm or the LMS. Explicitly, no off-line gradient estimation from repetitions of the data is required. Each iteration of the algorithm requires only the input vector and desired response.

The ODE Method

Motivation

Consider the adaptive form

$$\theta_n = \theta_{n-1} + \gamma_n H(\theta_{n-1}, X_n) \tag{14}$$

The nonlinear time-varying mapping from X_n to $\hat{\theta}_n$ makes analysis difficult however, the general application of this form suggests that a universal method of convergence analysis is a worthwhile goal. To this end a more precise definition of the following is required:

- Nature of the gain γ_n
- Nature of the state vector X_n
- Conditions on the vector field $H(\theta_{n-1}, X_n)$

Firstly, a heuristic argument for an associated differential equation will be given.

The Associated Differential Equation

Consider the following approximations where γ_n is replaced by a small constant γ:

$$\begin{aligned}
\theta_{n+N} &= \theta_n + \gamma \sum_{i=0}^{N-1} H(\theta_{n+i}, X_{n+i+1}) \\
&\approx \theta_n + \gamma \sum_{i=0}^{N-1} H(\theta_n, X_{n+i+1}) \\
&= \theta_n + (N\gamma) \cdot \frac{1}{N} \sum_{i=0}^{N-1} H(\theta_n, X_{n+i+1}) \\
&\approx \theta_n + N\gamma h(\theta_n)
\end{aligned} \tag{15}$$

The Associated Differential Equation

The approximations are justified as follows:

1. The function $\theta \to H(\theta, X)$ is regular
2. θ_{n+i} in N previous steps belong to a small neigbourhood of values
3. In the transversal equaliser state X_{n+i} is stationary
4. The mean vector field h is justified provided N is "large" enough

Points 2 and 4 present a conflict that is resolved by the need for small γ.

The Associated Differential Equation

The final expression has the form of a difference equation with discretisation $\theta_0, \theta_N, \theta_{2N} \ldots$. Another interpretation of this equation is a discrete version of the continuous time ODE:

$$\dot{\theta} = h(\theta) \quad \theta(0) = z \tag{16}$$

with solution

$$\theta(t) \quad \text{or} \quad \theta(z,t), \quad t \geq 0 \tag{17}$$

Regularity in terms of locally Lipschitz h then permits association of the ODE (assumption 1).

The Associated Differential Equation

Expressing (15) in the form of (16) yields

$$\theta_{n+1} = \theta_n + \gamma h(\theta_n) \tag{18}$$

with

$$\theta_n = \theta(t_n) \quad \text{with} \quad t_n = n\gamma \tag{19}$$

Note that if a small variable step size is used then shifting γ_n inside the summation then the concept of time is replaced

$$\theta_n = \theta(t_n) \quad \text{with} \quad t_n = \sum_{i=1}^{n} \gamma_i \tag{20}$$

It then seems reasonable that asymptotic properties of (14) can be studied in terms of (16).

The State Vector

The state vector (X_n) must be asymptotically stationary and regular in the limit of θ. Why?

Conditional Linear Dynamics

Algorithms for system identification often use models incorporating state dependent feedback. Stationarity is therefore tied to model stability. For example, a model based on rational stable transfer functions permits the following state representation[5]

$$\begin{aligned} \theta_n &= \theta_{n-1} + \gamma_n H\left(\theta_{n-1}, X_n\right) \\ X_n &= A\left(\theta_{n-1}\right) X_{n-1} + B\left(\theta_{n-1}\right) W_n \end{aligned} \tag{21}$$

where W_n is a stationary sequence of independent variables.

[5] This form was introduced in Lennart Ljung, 'Analysis of recursive stochastic algorithms', *IEEE transactions on automatic control* 22.4 (1977): 551–575 eq (1) and (2).

The State Vector

(21) is a special case of a more general representation of the state vector in which X_n is a functional of a Markov chain controlled by θ_n[6]:

$$X_n = f(\xi_n) \tag{22}$$

where (ξ_n) has a conditional distribution of the form

$$P(\xi_n \in G \mid \xi_{n-1}, \xi_{n-2}, \ldots; \theta_{n-1}, \theta_{n-2}, \ldots) = \int_G \pi_{\theta_{n-1}}(\xi_{n-1}, dx) \tag{23}$$

where θ is fixed and $\pi_\theta(\xi, dx)$ is the transition probability. Furthermore, (X_n) is asymptotically stationary and regular in θ at its limit.

[6] Harold J Kushner and Adam Shwartz, 'An invariant measure approach to the convergence of stochastic approximations with state dependent noise', *SIAM Journal on Control and Optimization* 22.1 (1984): 13–27.

The State Vector

Therefore in the Markov state model it is the transition probabilities $\pi_\theta^k(\xi_n, dx)$ that must display regularity and asymptotic convergence i.e.

- For fixed θ in the algorithm domain, conditional distributions of the Markov chain converge to a unique invariant probability:

$$\pi_\theta^k(\xi_n, .) \to \mu_\theta \tag{24}$$

- Function $\theta \to \mu_\theta$ is regular

The State Vector

The following proposition[7] underlies the importance of the Markov representation:

1. The following transformations preserve the Markov representation of X_n:

$$Y_n = g(X_n) \tag{25}$$

   where g is a suitably regular function and

$$Z_n = (X_n, \ldots, X_{n-p}) \tag{26}$$

   where p is a fixed integer.

2. The following transformation preserves the Markov representations of X_n and a_n which are controlled by the same extended state ξ_n:

$$U_n = (X_n, a_n) \tag{27}$$

[7] Benveniste, Métivier and Priouret page 27

Nature of the Gain

The first condition is on how fast the gain tends to zero:

$$\sum_n \gamma_n^\alpha < \infty \text{ for some } \alpha > 1 \tag{28}$$

The second condition is required for the alorithm to move the estimate to the desired limit:

$$\sum_n \gamma_n = +\infty \tag{29}$$

The Mean Vector Field

As alluded to earlier, for small γ it is expected that parameter estimates vary slowly compared to the state. This means we need to assume smoothness of $H(\theta, X_n)$. Therefore we define the mean vector field

$$h(\theta) := \lim_{n \to \infty} E_\theta \left(H(\theta, X_n) \right) \tag{30}$$

The expectation is over the distribution of X_n for a fixed value of θ. This condition is key to associating the ODE with the general adaptive form.

Summary of Conditions

A summary of conditions:

- General adaptive form (14) (omitting the complementary term)
- State vector (X_n) is stationary and indpendent of θ
- Estimation of fixed parameters requires decreasing gain
- Existence and regularity of the mean vector field $h(\theta)$ defined by (30)

We can now state some fundamental theorems for the study of convergence[8]

[8] See Ljung Appendices I,II and III for proofs

Finite Horizon

Theorem 1

Assuming all conditions are satisfied inside the cylinder of diameter $\eta > 0$ containing the trajectory $(\theta(t))_{0 \leq t \leq T}$ of the ODE, let

$$\gamma := \max_{n:t_n \leq T} \gamma_n \tag{31}$$

Then, for fixed $\varepsilon > 0$ and γ sufficiently small, we have

$$P\left\{ \max_{n:t_n \leq T} \|\theta_n - \theta(t_n)\| > \varepsilon \right\} \leq C(\gamma, T) \tag{32}$$

where, for fixed $T < \infty$, constants $C(\gamma, T)$ tend to zero as γ tends to 0.

Infinite Horizon

Convergence in the case $n \to \infty$ requires additional conditions.

Decreasing gain

The gain must sastify the following:

$$\sum_n \gamma_n^\alpha < \infty \text{ for some } \alpha > 1, \quad \sum_n \gamma_n = +\infty \tag{33}$$

1. Variance of the parameters tends to zero
2. Ensure $t_n \to \infty$

Infinite Horizon

Domain of attraction

This condition says that the ODE has an attractor θ_* with domain of attraction is D_*.

- ODE is asymptotically stable
- θ_* may be a point or limit cycle
- D_* is the set of all intial conditions z where $\theta(z, t)$ converge to θ_*

Infinite Horizon

Theorem 2

Assuming all conditions are satisfied in D_* and suppose that algorithm (14) is initialised with $\theta_0 = z \in Q$, where Q is a compact subset of D_* and $\xi_0 = \xi$. Then

$$
\begin{aligned}
&(i) \qquad P\left\{\lim_{n\to\infty}\theta_n = \theta_*\right\} \geq 1 - C(\alpha, Q, |\xi|)\sum_n \gamma_n^\alpha \\
&(ii) \qquad P\left\{\max_n \|\theta_n - \theta(z, t_n)\| > \varepsilon\right\} < C(\alpha, Q, |\xi|)\sum_n \gamma_n^\alpha
\end{aligned}
\tag{34}
$$

for any $\varepsilon > 0$

C here is a constant that depends on α, compact subset Q and norm of the initial condition and allows some degree of control over the error.

Infinite Horizon

It remains to be discussed that there are risks of divergent θ_n. There are two mechanisms

Behaviour of state

The output of the state Markov chain π_θ is transient instead of recurrent. This can happen despite θ within a stable-state domain. $\theta \to \infty$ is indicative of this behaviour.

Infinite Horizon

Boundedness Condition

Conversely θ_n may leave the domain of attraction because:

- $\gamma_n H(\theta_{n-1}, X_n)$ may be too large even for small γ
- Variance of $H(\theta_{n-1}, X_n)$ may increase to the extent that a "random walk" effect to infinity takes place

It can only be verified for a specfic algorithm that

$$P\{\theta_n \in Q \text{ infinitely often } \} = 1 \tag{35}$$

.

Summary

A summary of all the theorems stated might be given as follows: after the transient phase, the behaviour of algorithm (14) is represented to a first approximation by that of the ODE (16) and (17).

Analysis of Adaptive Algorithms

Recap

Form (14) can be studied and analysed in terms of differential equation (16). A more detailed summary of the theoretical results might be the following:

1. θ_n can only converge to stable stationary points of the differential equation
2. If $\theta(t)$ belongs to a domain of attraction of the stable stationary point θ^* of (16) infinitely often a.s. then θ_n converges to θ^* a.s. as $n \to \infty$
3. Trajectories of (16) can be described as asymptotic paths of estimates θ_n generated by the update equation.

Given a set of reasonably weak conditions.

Analysis of Adaptive Algorithms: Stages of Adaptive Analysis

Stage 1

Express the algorithm in the general form and verify that the theory is applicable

1. Identify H, ε_n, γ_n, X_n, ξ_n
2. Is the gain decreasing or constant?
3. Identify θ_n for which ξ_n blows up
4. Does ξ_n have unique stationary asymptotic beaviour: ξ_n and ξ_{n+N} independence

Analysis of Adaptive Algorithms: Stages of Adaptive Analysis

Stage 2:

1. Calculation of the mean vector field $h(\theta)$

Analysis of Adaptive Algorithms: Stages of Adaptive Analysis

Stage 3:

Study the ODE

1. Classical analysis of a differential equation
2. Study of attractors and their domains of attraction
3. What are the trajectories of the ODE?
4. Where possible choose a suitable Lyapunov function J
5. Consider local minima of J, the potential from which $h(\theta)$ is derived and their domains of attaction.

Analysis of the Tranversal Equaliser

Stage 1: Expression of the Algorithm in the General Form

- γ_n is decreasing for a learning phase
- X_n and H were identified in (11) and restated as follows:

$$\begin{aligned} X_n &= \begin{pmatrix} a_n \\ Y_n \end{pmatrix} \\ H(\theta, X_n) &= Y_n \left[a_n - Y_n^T \cdot \theta_{n-1} \right] \end{aligned} \tag{36}$$

- (y_n) is generated from an ARMA process and therefore (Y_n) is asymptotically stationary. Similarly for (X_n)
- H is regular
- Algorithm behaviour therefore given by ODE

Analysis of the Transversal Equaliser

Stage 2: Calculation of the ODE

- Evaluate the mean vector field:

$$h(\theta) = E\left(H\left(\theta, X_n\right)\right) \tag{37}$$

- Expectation is with respect to the stationary asymptotic distribution of X_n.

Analysis of the Transversal Equaliser

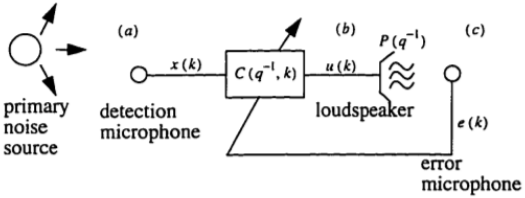
Stage 3: Analysis of the ODE

- A suitable Lyapunov function is $V(\theta) = E\left(e_n^2(\theta)\right)$ where $e_n(\theta) = a_n - Y_n^T \theta$
- Satisfies $h(\theta) = -\frac{d}{d\theta} V(\theta)$
- Trajectories of the ODE are the lines of steepest descent of the mean square error between true message and the output of the equaliser
- V is strictly decreasing function but bounded below
- θ must therefore converge to the attractor θ_*

Applications II

Description

The following is a typical broadband feedforward ANC system:



The objective in ANC is to cancel an acoustic noise source at a position (c). An error signal is derived from this point and is used to drive adaption of a filter. The noise source is detected at (a), filtered and used to drive a loudspeaker to generate an “anti-noise”.

Description

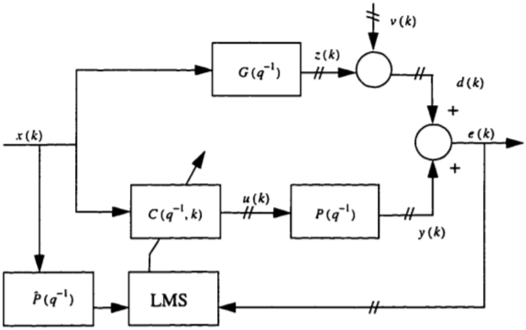
Two elements to the ANC problem:

- Prediction
- Equalisation

Prediction of the sound pressure level of the acoustic noise is used to compensate for inherent time delay in actuation. Equalisation provides compensation for non-linearities and unmodelled dynamics in the actuator transfer function

Problem Description

Feedforward control is achieved using a variant of the LMS algorithm called filtered-X LMS that shifts the secondary path transfer function filter to the plant input[9]:



[9] See Bernard Widrow and Samuel D Stearns, 'Adaptive Signal Processing Prentice-Hall', *Englewood Cliffs, NJ* (1985) chapter 11.

Description

Symbols are defined as follows:

- $P(q^{-1})$ is the speaker to error microphone transfer function
- $C(q^{-1})$ is the controller, a transversal (all-zero) filter
- $G(q^{-1})$ is the transfer function capturing dynamics of the disturbance to the error microphone
- $v(k)$ is a measurement noise (also known as "plant noise" and a source of stochasticity)

Expression in the general form

Guided by the feedforward control scheme and general adaptive form of (14) we develop a the filtered-X LMS algorithm.

Controller

A transveral (all-zero) model is used for the controller such that

$$C\left(q^{-1}\right) = c_0 + c_1 q^{-1} + \ldots + c_N q^{-N} \tag{38}$$

At any given iteration of the algorithm the controller weights are given by the parameter vector of length $N+1$:

$$\hat{\theta}(k) = [c_0, c_1, \ldots, c_N]^T \tag{39}$$

Expression in the general form

Criterion and Error Function

Again we minimise a squared error criterion to derive the gradient approximation and therefore a search direction along which to update controller coefficients. From the control diagram we have

$$\begin{aligned} e(k) &= z(k) + v(k) + P\left(q^{-1}\right) u(k) \\ &= z(k) + v(k) + P\left(q^{-1}\right)\left(\phi^{T}(k)\hat{\theta}(k)\right) \end{aligned} \tag{40}$$

where

$$\phi(k) = [x(k), x(k-1), \ldots, x(k-N)]^{T} \tag{41}$$

is the state vector.

Expression in the general form

Gradient approximation

The objective of the algorithm is to minimise a quadratic error criterion, namely

$$J = \frac{1}{2}E[e^2(k)] \tag{42}$$

Taking the derivative with respective to each parameter gives

$$\frac{\partial J}{\partial \theta_i} = E\left[e(k)\frac{\partial e(k)}{\partial \theta_i}\right] \tag{43}$$

Differentiating (40) with respect to each coefficient:

$$\frac{\partial e(k)}{\partial \hat{\theta}_i} = \sum_{j=0}^{N} p_j \phi(k-i-j) \tag{44}$$

Expression in the general form

Filtered-X LMS Algorithm

The cofficients may then be updated by a proportion of the negative instantaneous value of the gradient[10]:

$$\hat{\theta}(k+1) = \hat{\theta}(k) - \gamma(k)\left[\hat{P}\left(q^{-1}\right)\phi(k)\right]e(k) \tag{45}$$

where γ is a positive adaption gain.

[10] This derivation is due to Stephen Elliott, IANM Stothers and Philip Nelson, 'A multiple error LMS algorithm and its application to the active control of sound and vibration', *IEEE Transactions on Acoustics, Speech, and Signal Processing* 35.10 (1987): 1423–1434 but is equivalent to the filtered-X LMS algorithm in Widrow and Stearns

Recap

Recall the basic recipe for convergence analysis:

1. Express the algorithm in the general form
2. Freeze the model paramter estimate $\hat{\theta}$, compute the state vector and prediction error and evaluate the mean vector field of the associated ODE.
3. Study the stability of the associated ODE defined by the mean vector field.

Single Input Single Output Analysis

Step 1: Regularity conditions

The algorithm is defined by the combination of a small gain sequence and the update term $H(\hat{\theta}(k), \phi(k))$ in (45). The smoothness of H is satisfied as follows:

- No specific assumption on the error term
- $H(\hat{\theta}(k), \phi(k))$ is Lipschitz continuous in θ and ϕ
- H and its derivatives must not increase rapidly with ϕ and e

for all θ within a compact subset of the stability region[11]. Regularity conditions are therefore satisfied in this example.

[11] This is simply a projection mechanism to ensure that all estimates are confined to the stability region and is an implicit assumption of the method

Single Input Single Output Analysis

Step 2: Freeze the trajectories

Next freeze the trajectories of the state and estimation error:

$$\bar{\phi}(k,\theta) = \phi(k) \tag{46}$$

$$\bar{e}(k) = d(k) + P\left(q^{-1}\right)\left(\phi^T(k)\theta\right) \tag{47}$$

Single Input Single Output Analysis

Step 2: Averge update term

Mean vector field is given by:

$$\begin{aligned} f(\theta) &= -E\left[\left(\hat{P}\left(q^{-1}\right)\bar{\phi}(k,\theta)\right)\bar{e}(k,\theta)\right] \\ &= -E\left[\hat{P}\left(q^{-1}\right)\phi(k)\right]\left[z(k)+v(k)+P\left(q^{-1}\right)\left(\phi^T(k)\theta\right)\right] \\ &= -E\left[\left(\hat{P}\left(q^{-1}\right)\phi(k)\right)z(k)\right] \\ &\quad - E\left[\left(\hat{P}\left(q^{-1}\right)\phi(k)\right)P\left(q^{-1}\right)\left(\phi^T(k)\theta\right)\right] \end{aligned} \tag{48}$$

where we have imposed the requirement that $x(k)$ and $v(k)$ are uncorrelated.

Single Input Single Output Analysis

Step 2: Define the associated ODE

The associated ODE is then:

$$\begin{aligned}\frac{d}{d\tau}\theta(\tau) = &-E\left[\left(\hat{P}\left(q^{-1}\right)\phi(k)\right)P\left(q^{-1}\right)\phi^T(k)\right]\theta(\tau)\\ &-E\left[\left(\hat{P}\left(q^{-1}\right)\phi(k)\right)z(k)\right]\end{aligned} \tag{49}$$

and has a unique equilibrium point (no change in the update)

$$\begin{aligned}\theta^* = &-E\left[\left(\hat{P}\left(q^{-1}\right)\phi(k)\right)P\left(q^{-1}\right)\phi^T(k)\right]^{-1}\times\\ &E\left[\left(\hat{P}\left(q^{-1}\right)\phi(k)\right)z(k)\right]\end{aligned} \tag{50}$$

for non-singular $E\left[\left(\hat{P}\left(q^{-1}\right)\phi(k)\right)P\left(q^{-1}\right)\phi^T(k)\right]$

Single Input Single Output Analysis

Step 3: Analysis of the ODE

Now that the ODE is associated with the algorithm its asymptotics should follow the asymptotics of the ODE trajectories. ODE asymptotics are typically expressed in terms of stability. This can be demonstrated with:

- Numerical simulation
- A suitable Lyapunov function

We shall find a Lyapunov function since it is a necessary and sufficient condition for stability for this ODE.

Single Input Single Output Analysis

Step 3: Analysis of the ODE (continued)

In brief:

- Define a Lyapunov function $V(x) \geq 0$ for all x such that when evaluated along solutions to the ODE:

$$\begin{aligned} \frac{d}{d\tau}V(x(\tau)) &= \frac{d}{dx}V(x(\tau)) \cdot \frac{d}{d\tau}x(\tau) \\ &= V'(x(\tau))f(x(\tau)) \leq 0 \text{ for all } x(\tau) \end{aligned} \tag{51}$$

and

$$\frac{d}{d\tau}V(x(\tau)) = 0 \Rightarrow x(\tau) \in D_c \tag{52}$$

Single Input Single Output Analysis

Step 3: Analysis of the ODE (continued)

Consider the following system:

$$\frac{d}{dt}x(t) = Ax(t) \tag{53}$$

and Lyapunov function:

$$V(x) = x^T(t)x(t) \tag{54}$$

Then if eigenvalues of A are in the open left half of the complex plane for system the equilibrium point stable.

Single Input Single Output Analysis

Step 3: Analysis of the ODE (continued)

The ODE is a linear time-invariant differential equation and its stability is determined by the eigenvalues of the following $N \times N$ matrix:

$$A \equiv -E\left[\left(\hat{P}\left(q^{-1}\right)\phi(k)\right)P\left(q^{-1}\right)\phi(k)\right] \tag{55}$$

The significance of A is that it is a generalisation of the classical reference signal autocorrelation matrix[12].

[12] See Widrow and Stearns page 20

Single Input Single Output Analysis

Step 3: Analysis of the ODE (continued)

Let L be a non-zero $(N \times 1)$ vector. It can be shown that[13]:

$$\begin{aligned} L^T A L &= -E\left[L^T \hat{P}\left(q^{-1}\right)\phi(k) P\left(q^{-1}\right)\phi^T(k) L\right] \\ &= -E\left[\hat{P}\left(q^{-1}\right)\left(L^T\phi(k)\right) P\left(q^{-1}\right)\left(L^T\phi(k)\right)\right] \\ &= -\frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{Re}\left[\hat{P}^*\left(e^{-j\omega}\right) P\left(e^{-j\omega}\right) S_\gamma\left(e^{-j\omega}\right)\right] d\omega \end{aligned} \tag{56}$$

where $S_\gamma\left(e^{-j\omega}\right)$ is the power spectral density of process $L^T\phi(k)$ and noting that $\phi(k)$ contains a history of reference samples which is assumed to be stationary.

[13] See Angela Kuo Wang, 'Adaptive algorithms for active noise control.' (1997)

Single Input Single Output Analysis

Step 3: Boundedness Condition

Under the assumption that transfer function $\hat{P}^*\left(q^{-1}\right) P\left(q^{-1}\right)$ is strictly positive real then it is also stable i.e.

$$\operatorname{Re}\left[\hat{P}^*\left(e^{-j\omega}\right) P\left(e^{-j\omega}\right)\right] > 0 \quad \text{for all } -\pi \leq \omega \leq \pi \tag{57}$$

It follows:

$$\frac{1}{2\pi}\int_{-\pi}^{\pi} \operatorname{Re}\left[\hat{P}^*\left(e^{-j\omega}\right) P\left(e^{-j\omega}\right) S_\gamma\left(e^{-j\omega}\right)\right] d\omega > 0 \tag{58}$$

For $S_\gamma\left(e^{-j\omega}\right) > 0$ for all ω. Therefore $L^T A L < 0$ for $L \neq 0$.

Single Input Single Output Analysis

Summary of Convergence Analysis

Boundedness of $\hat{\theta}(k)$ is required by the ODE theorem[14]. However interestingly for linear problems this boundedness condition can be omitted for linear algorithms provided some conditions are placed on the step size and average update direction[15].

[14] Ljung

[15] Michel Metivier and Pierre Priouret, 'Applications of a Kushner and Clark lemma to general classes of stochastic algorithms', *IEEE Transactions on Information Theory* 30.2 (1984): 140–151

Single Input Single Output Analysis

Analysis of the ODE

Summary of assumptions:

- $\hat{P}^* \left(q^{-1}\right) P \left(q^{-1}\right)$ is SPR
- $\gamma(k) \to 0$ as $k \to \infty$
- Persistence of excitation of the reference signal $x(k)$
- Regularity conditions[16]

Then the filtered-X LMS algorithm converges with probability 1 to the unique stable equilibrium point θ^*.

[16] See Lennart Ljung and Torsten Söderström, *Theory and practice of recursive identification* (MIT press, 1983) page xx