T-61.3040 Statistical Signal Modeling Lendasse, Miche, Eirola
Exercise 11, December 8, 2011, Solutions

1. (a) The z-transform of the difference equation of the MA process is

$$X(z) = (1 + \alpha z^{-1})V(z)$$

so that the transfer function of the process is

$$H(z) = 1 + \alpha z^{-1}$$

and the spectral density or the power spectrum is

$$\begin{aligned} P_x(z) &= H(z)H^*(1/z^*)P_v(z) \\ &= (1 + \alpha z^{-1})(1 + \alpha^* z)P_v(z) \\ &= (1 + |\alpha|^2 + \alpha z^{-1} + \alpha^* z)P_v(z) \end{aligned}$$

Because the noise $v(n)$ is white and has unit variance, $P_v(z) = 1$, and the spectral density of the MA process is

$$\begin{aligned} P_x(e^{j\omega}) &= 1 + |\alpha|^2 + \alpha e^{-j\omega} + \alpha^* e^{j\omega} \\ &= 1 + |\alpha|^2 + |\alpha| e^{j\phi} e^{-j\omega} + |\alpha| e^{-j\phi} e^{j\omega} \\ &= 1 + |\alpha|^2 + 2|\alpha| \cos(\omega - \phi) \end{aligned}$$

where ϕ is the argument (phase) of α. The maximum value is $(1 + |\alpha|)^2$ and the minimum value $(1 - |\alpha|)^2$, resulting the upper bound for the condition number of the correlation matrix

$$\chi \leq \frac{(1 + |\alpha|)^2}{(1 - |\alpha|)^2}$$

(b) Now the transfer function is

$$H(z) = \frac{1}{1 - \alpha z^{-1}}, \quad |\alpha| < 1$$

resulting in the spectral density

$$P_x(z) = \frac{1}{1 - \alpha z^{-1}} \frac{1}{(1 - \alpha z^*)^*} = \frac{1}{1 + |\alpha|^2 - \alpha z^{-1} - \alpha^* z}$$

so that

$$P_x(e^{j\omega}) = \frac{1}{1 + |\alpha|^2 - 2|\alpha| \cos(\omega - \phi)}$$

which yields the bound for the condition number

$$\chi \leq \frac{1}{(1 - |\alpha|)^2} \Big/ \frac{1}{(1 + |\alpha|)^2} = \frac{(1 + |\alpha|)^2}{(1 - |\alpha|)^2}$$

NB:

- If the condition number is large, one says that the matrix is "ill-conditioned". In that case one has difficulties for example in calculating the inverse of the matrix.

- Let x and y be arbitrary orthogonal vectors with unit length, A a matrix, χ the condition number of that matrix and θ the angle between the vectors Ax and Ay. Then we have $\chi = \min_{x,y} \cot(\theta/2)$. If the condition number χ of A is large, the minimum θ is small. In other words, under the mapping by an ill-conditioned matrix at least some orthogonal vectors become nearly parallel.
- By using the condition number, one can quantify the sensitivity to perturbations in initial values. For example in Wiener filtering the effect of relative perturbations ρ_R and ρ_r initial values (the autocorrelation matrix $\mathbf{R}_y$ and the cross correlation vector $\mathbf{r}_{dy}$) on the output of the filter is at most $\chi(\mathbf{R}_y)(\rho_R + \rho_r)$ in magnitude.

2. (a) The estimate of the target signal is $\hat{d}(n) = \sum_k w(k)x(n-k)$ and the error is defined as $e(n) = d(n) - \hat{d}(n)$. Let's calculate the derivatives of the function $E(|e(n)|)$ with respect to the parameters $w(k)$. The differentiation is carried out inside the expected value operator, resulting in

$$\frac{\partial \,\mathrm{E}(|e(n)|)}{\partial w(k)} = \mathrm{E}\left(\frac{\partial |e(n)|}{\partial w(k)}\right)$$

The derivative of the function $f(x) = |x|$ is 1 when $x > 0$ and -1 when $x < 0$. So

$$\mathrm{E}\left(\frac{\partial |e(n)|}{\partial w(k)}\right) = \mathrm{E}\left(\mathrm{sgn}(e(n))\frac{\partial e(n)}{\partial w(k)}\right) = \mathrm{E}(-\,\mathrm{sgn}(e(n))x(n-k))$$

because the partial derivative of the error is simply

$$\frac{\partial e(n)}{\partial w(k)} = \frac{\partial}{\partial w(k)}\left(d(n) - \sum_l w(l)x(n-l)\right) = -x(n-k)$$

The goal is to update each parameter $w(k)$ in the direction of the negative gradient (steepest descent), i.e.,

$$w_{n+1}(k) = w_n(k) - \mu\frac{\partial e(n)}{\partial w(k)} = w_n(k) + \mu\,\mathrm{E}(\mathrm{sgn}(e(n))x(n-k))$$

where μ is some step size. Or, written in vector form:

$$\mathbf{w}_{n+1} = \mathbf{w}_n + \mu\,\mathrm{E}(\mathrm{sgn}(e(n))\mathbf{x}(n))$$

(b) By replacing the expected value with with the pointwise estimate $\mathrm{sgn}(e(n))\mathbf{x}(n)$, one gets the update rule

$$\mathbf{w}_{n+1} = \mathbf{w}_n + \mu\,\mathrm{sgn}(e(n))\mathbf{x}(n)$$

This is known as the *sign-error* LMS algorithm (see Hayes, p. 523).

3. The error fuctional to be minimised is

$$\begin{aligned}\xi'(n) &= |e(n)|^2 + \beta \mathbf{w}_n^H \mathbf{w}_n = e(n)e^*(n) + \beta \mathbf{w}_n^H \mathbf{w}_n \\ &= (d(n) - \mathbf{w}_n^T \mathbf{x}(n))(d(n) - \mathbf{w}_n^T \mathbf{x}(n))^* + \beta \mathbf{w}_n^H \mathbf{w}_n\end{aligned}$$

with the gradient

$$\nabla_{\mathbf{w}^*} \xi'(n) = e(n) \nabla e^*(n) + \beta \mathbf{w}_n = -e(n)\mathbf{x}^*(n) + \beta \mathbf{w}_n$$

(a) The update step for the weight vector:

$$\begin{aligned}\mathbf{w}_{n+1} &= \mathbf{w}_n - \mu \nabla_{\mathbf{w}^*} \xi'(n) \\ &= \mathbf{w}_n + \mu e(n)\mathbf{x}^*(n) - \beta\mu \mathbf{w}_n \\ &= (1 - \beta\mu)\mathbf{w}_n + \mu e(n)\mathbf{x}^*(n)\end{aligned}$$

We notice that this is the same update step as in the Leaky LMS algorithm (Hayes, p. 522) if one sets $\beta = \gamma$.

(b) By inserting $e(n) = d(n) - \mathbf{w}_n^T \mathbf{x}(n)$ in the above expression and rearranging the terms containing $\mathbf{w}_n$ to be separate one obtains

$$\begin{aligned}\mathbf{w}_{n+1} &= (1 - \beta\mu)\mathbf{w}_n + \mu(d(n) - \mathbf{x}(n)^T \mathbf{w}_n)\mathbf{x}^*(n) \\ &= \mathbf{w}_n - \mu\beta \mathbf{w}_n + \mu d(n)\mathbf{x}^*(n) - \mu \mathbf{x}^*(n)\mathbf{x}(n)^T \mathbf{w}_n \\ &= [\mathbf{I} - \mu(\mathbf{x}^*(n)\mathbf{x}^T(n) + \beta \mathbf{I})]\mathbf{w}_n + \mu d(n)\mathbf{x}^*(n)\end{aligned}$$

By assuming $\mathbf{x}(n)$ and $\mathbf{w}_n$ are independent, one gets by taking expectations

$$\mathrm{E}(\mathbf{w}_{n+1}) = [\mathbf{I} - \mu(\mathbf{R}_x + \beta \mathbf{I})]\mathrm{E}(\mathbf{w}_n) + \mu \mathbf{r}_{dx}$$

This is nearly identical to the the original LMS. The only difference is that instead of $\mathbf{R}_x$, here we have $\mathbf{R}_x + \beta \mathbf{I}$. The convergence can thus be analysed just as that of the LMS algorithm (see Hayes pp. 500-503 and 506-508 for more information), if one replaces the eigenvalues λ_k of $\mathbf{R}_x$ with the eigenvalues $\lambda_k + \beta$ of the matrix in the analysis. This way one can bound the step length by

$$\mu < \frac{2}{\lambda_{\max} + \beta}$$

(c) The change of the weight vector

$$\mathrm{E}(\mathbf{w}_{n+1}) - \mathrm{E}(\mathbf{w}_n) = -\mu[(\mathbf{R}_x + \beta \mathbf{I})\,\mathrm{E}(\mathbf{w}_n) - \mathbf{r}_{dx}]$$

is zero on average when

$$\mathrm{E}(\mathbf{w}_n) = (\mathbf{R}_x + \beta \mathbf{I})^{-1} \mathbf{r}_{dx}$$

As with the LMS algorithm, the method will converge to this limit, provided the step size is chosen appropriately. Note that the limit is equivalent to solving the Wiener-Hopf equations $\mathbf{R}_x \mathbf{w}_n = \mathbf{r}_{dx}$ with applied *Tikhonov regularisation* (also known as *ridge regression*), i.e., $(\mathbf{R}_x + \beta \mathbf{I})\mathbf{w}_n = \mathbf{r}_{dx}$.

4. (a) Because $x(n)$ is always normally distributed with zero mean, the undonditional and conditional means are both zero. The undonditional variance is given by the iteration formula

$$\begin{aligned}\mathrm{Var}(x(n)) =& \mathrm{E}[\mathrm{Var}(x(n)|x(n-1))] + \mathrm{Var}(\mathrm{E}[x(n)|x(n-1)]) \\ =& P(x(n-1) < 0)\,\mathrm{Var}(x(n)|x(n-1) < 0) \\ &+ P(x(n-1) > 0)\,\mathrm{Var}(x(n)|x(n-1) > 0) + \mathrm{Var}(0) \\ =& 0.5 \cdot 1 + 0.5 \cdot 2 + 0 = 1.5\end{aligned}$$

The conditional variance $\mathrm{Var}(x(n)|x(n-1))$ is directly $0.5 \cdot (\mathrm{sgn}(x(n-1)) + 3)$. This is not a constant but depends on the observation $x(n-1)$.

(b) A process is WSS if the following requirements are fulfilled:

1. The mean of the process is constant.
2. The autocorrelations $r_x(k, l)$ depend only on the difference $k - l$.
3. The variance is finite.

The mean of the process $x(n)$ is zero for all n, and the variance $\mathrm{E}[x^2(n)] = r_x(0) = 1.5 < \infty$ for all n.

Let's still evaluate the autocorrelations $r_x(n, n-k)$ for lags k that are different from zero. It is enough to consider positive values for k as the autocorrelation has conjugate symmetry w.r.t. the change of the order of its arguments.

If we define $g(u) = \sqrt{0.5 \cdot (\mathrm{sgn}(u) + 3)}$, then we can write the process as $x(n) = g(x(n-1))v(n)$, where $v(n)$ is normally distributed white noise with unit variance and zero mean. Note that all $v(n)$ are mutually independent, and $g(x(k))$ and $v(n)$ are independent for all n and k except $k = n$.

Then for the autocorrelations:

$$\begin{aligned}r_x(n+k, n) &= \mathrm{E}[x(n+k)x^*(n)] \\ &= \mathrm{E}[g(x(n+k-1))v(n+k)g(x(n-1))v^*(n)]\end{aligned}$$

Here there are two separate cases to consider. Either $k > 1$, in which case all terms in the product are independent, and:

$$r_x(n+k, n) = \mathrm{E}[g(x(n+k-1))]\,\mathrm{E}[g(x(n-1))]\,\underbrace{\mathrm{E}[v(n+k)]}_{=0}\,\underbrace{\mathrm{E}[v^*(n)]}_{=0} = 0$$

or $k = 1$, and

$$\begin{aligned}r_x(n+1, n) &= \mathrm{E}[g(x(n+1-1))v(n+1)g(x(n-1))v^*(n)] \\ &= \underbrace{\mathrm{E}[g(x(n))v^*(n)]}_{=\text{something finite}}\,\mathrm{E}[g(x(n-1))]\,\underbrace{\mathrm{E}[v(n+1)]}_{=0} = 0\end{aligned}$$

All the autocorrelations for non-zero lag are then 0. In particular, they are independent of n, satisfying requirement 2 above. Hence, the process $x(n)$ is WSS.

5. (a) For the 2×2 autocorrelation matrix we need $r_x(0) = 1$ and $r_x(1) = a$, so

$$\mathbf{R}_x = \begin{bmatrix} 1 & a \\ a & 1 \end{bmatrix}.$$

From the eigenvalue equation $\mathbf{R}_x\mathbf{e} = \lambda\mathbf{e}$ we get

$$\begin{aligned} e_1 + ae_2 &= \lambda e_1 \\ ae_1 + e_2 &= \lambda e_2, \end{aligned}$$

that is

$$\begin{aligned} (1-\lambda)e_1 &= -ae_2 \\ (1-\lambda)e_2 &= -ae_1. \end{aligned}$$

Inserting $1 - \lambda = -a$ we get $e_1 = e_2$, so $\lambda_1 = 1 + a$ is one eigenvalue and the corresponding eigenvector is $\mathbf{e} = [1\ 1]^T$. Similarly $1 - \lambda{=}a$ gives $e_1 = -e_2$ provides $\lambda_2 = 1 - a$ as the other eigenvalue, with the eigenvector $\mathbf{e} = [1\ -1]^T$.

(b) According to the hint, it is sufficient to find the maximum value of the power spectrum, as

$$\lambda_{max} \to \max_{\omega} P_x(e^{j\omega})$$

The power spectrum is

$$\begin{aligned} P_x(e^{j\omega}) &= \sum_{k=-\infty}^{\infty} e^{-j\omega k} a^{|k|} = \sum_{k=0}^{\infty} e^{j\omega k} a^k + \sum_{k=0}^{\infty} e^{-j\omega k} a^k - 1 \\ &= \frac{1}{1-ae^{-j\omega}} + \frac{1}{1-ae^{j\omega}} - 1 \\ &= \frac{1 - ae^{j\omega} + 1 - ae^{-j\omega} - (1-ae^{j\omega})(1-ae^{-j\omega})}{(1-ae^{j\omega})(1-ae^{-j\omega})} \\ &= \frac{1 - ae^{j\omega} + 1 - ae^{-j\omega} - 1 + ae^{j\omega} + ae^{-j\omega} - a^2}{1 - ae^{j\omega} - ae^{-j\omega} + a^2} \\ &= \frac{1-a^2}{1 + a^2 - a(e^{j\omega} + e^{-j\omega})} \\ &= \frac{1-a^2}{1 + a^2 - 2a\cos\omega} \end{aligned}$$

If $a > 0$, the maximum occurs at the frequency $\omega = 0$. And if $a < 0$, the maximum is at the frequency $\omega = \pi$. In either case, the maximum value is

$$\frac{1-a^2}{1+a^2-2|a|} = \frac{(1+|a|)(1-|a|)}{(1-|a|)^2} = \frac{1+|a|}{1-|a|}$$

which is the asymptotic value of the largest eigenvalue of $\mathbf{R}_x$ as $p \to \infty$.

(c) For the LMS algorithm, the convergence is limited by an upper bound for the step size μ given by $2/\lambda_{\max}$ Inerting the value calculated in part (b) then gives the upper bound

$$2\frac{1-|a|}{1+|a|}$$

As $|a| \to 1$, the upper bound is close to 0, and the step size must be chosen to be very small.