Errata and Notes for

Statistical and Adaptive Signal Processing

Dimitris G. Manolakis, Vinay K. Ingle, and Stephen M. Kogon,
Artech Housre, Inc., © 2005, ISBN 1580536107.

James McNames

March 15, 2006

Errata are denoted with a $*$ and my notes and comments are denoted with a $\bullet$. This list was a joint effort of me and the students who were enrolled in my course, *Statistical Signal Processing*, during fall term of 2005 at Portland State University.

Chapter 2

$*$ **Page 42:** "band-imited" should be "band-limited" in the Sampling Theorem.

$*$ **Page 43:** The text states that the ROC of the z transform for causal signals includes $|z| < \infty$ and for anti-causal signals includes $0 < |z|$. These extreme values should be included for causal and anti-causal signals.

$*$ **Page 54:** (Second paragraph.) Technically, the unit sample sequence is an energy signal that has a constant energy spectrum, not "constant power spectrum values".

$*$ **Page 56:** There is a great deal of confusion in the text about maximum phase systems. The second paragraph on pg. 56 incorrectly states that a PZ system is maximum phase if it's poles and zeros are outside the unit circle. If we require that the system be stable, this means the system must also be anticausal. Furthermore, a maximum phase system can be obtained from a minimum phase system with the same magnitude response by $H_{\max}(z) = H_{\min}(z^{-1})$. All three of these statements are incorrect. A PZ system is maximum phase if it is causal, stable, and has all of its zeros outside of the unit circle. In order to be causal and stable, the poles are inside the unit circle. If the minimum phase system is given by $H_{\min}(z) = \frac{B_{\min}(z)}{A(z)}$, then the transformation to the maximum phase system is

$$H_{\max}(z) = \frac{z^{-Q} B^*_{\min}(z^{-*})}{A(z)} \tag{1}$$

where Q is the order of the polynomial $B_{\min}(z)$.

- **Page 59:** The solution to example 2.4.3 suggests the transformation

$$(1 - re^{j\theta} z^{-1}) \rightarrow r(1 - \frac{1}{r} e^{j\theta} z^{-1}). \tag{2}$$

Note that this requires that the zeros of $H_{\text{min}}(z)$ to occur in complex conjugate reciprocals in order for this to work. For example, consider if we apply the transformation to a minimum phase system with one complex zero.

$$\begin{aligned} H_{\text{min}}(z) &= (1 - z_k z^{-1}) \\ z_k &\triangleq re^{j\theta} \\ H_{\text{max}}(z) &= \frac{z^{-Q} B^*_{\text{min}}(z^{-*})}{A(z)} \\ &= z^{-1}(1 - z_k^* z) \\ &= -z_k(1 - z_k^{-*} z^{-1}) \\ &= -re^{j\theta} \left(1 - \frac{1}{r} e^{j\theta} z^{-1}\right) \end{aligned} \tag{3}$$

This differs by a factor of $-e^{j\theta}$ from the transformation listed in the book. However, this factor cancels out when we consider a pair of zeros that form complex conjugate pairs. The transformation in that case becomes

$$\begin{aligned} H_{\text{min}}(z) &= (1 - z_k z^{-1})(1 - z_k^* z^{-1}) \\ H_{\text{max}}(z) &= \left[-z_k(1 - z_k^{-*} z^{-1})\right] \left[-z_k^*(1 - z_k^{-1} z^{-1})\right] \\ &= r^2 \left(1 - \frac{1}{r} e^{j\theta} z^{-1}\right) \left(1 - \frac{1}{r} e^{-j\theta} z^{-1}\right) \end{aligned} \tag{4}$$

which is consistent with the transformation listed in the book.

∗ **Page 62:** The energy delays shown in the bottom plot of Figure 2.14 are incorrect. It appears that the plot shows the cumulative sum of the impulse responses, instead of the cumulative sum of the impulse response energies.

Chapter 3

- **Page 77:** The numbers that summarize key aspects of density functions are typically called "point statistics." I think this term is used more often than "statistical averages" or "moments". Although averages and moments are arguably the most common point statistics, there are some popular point statistics that are neither of these such as the median, modes, and percentiles of densities.

- **Page 78–79:** The image of two different types of Kurtosis in figure 3.2d is misleading. Kurtosis is mostly a property of how heavy the tails of the distribution are, not how pointed or flat the density is near the center as illustrated in this figure. Densities with heavy tails have a positive kurtosis and those with light or short tails have negative kurtosis. Similar comments apply to the description of kurtosis on page 79.

- **Page 79:** In the last sentence the text states "The *characteristic function* of a random variable $x(\zeta)$ is defined by ..." Strictly speaking, I think that should be described as the characteristic function of a *density*, not a random variable. This is a very minor point.

- **Page 82:** The first two sentences in the last paragraph of the section, "Normal random variable" are redundant with the prior paragraph.

* **Page 83:** The text states that because the Cauchy random variable does not have a finite variance, "the moment generating function does not exist, in general." This appears to contradict the statement that "the characteristic function always exists" on page 80, which is a special case of the moment generating function when $s = j\xi$. It seems to me that the moment generating function does exist for the Cauchy random variable (or at least there is a region of convergence with finite area that includes the imaginary s axis in the s domain), but the moment generating function is not analytic. That is, not all derivatives of the moment generating function $\overline{\Phi}_x(s)$ exist for the Cauchy random variable.

* **Page 86:** The matrix $\boldsymbol{R_x}$ is called the "correlation matrix" instead of the "autocorrelation matrix" at the end of the first paragraph.

- **Page 106:** The following statement at the end of the second paragraph is a bit misleading, "...because all realizations of a stationary random process exist for all time; that is, they are power signals." This may lead the reader to believe that all stationary random processes are power signals. While this is generally true in practice, it is not true in general. For example an IID sequence drawn from a Cauchy distribution is not a power signal, but it is a stationary random process. These statements would be equivalent if Definition 3.4 on page 102 also required that the variance be finite, as well is constant. Is an infinite variance constant? I think it's debatable.

* **Page 111:** In the middle of this page there is an equation $r_x(n_1, n_2) = \mathrm{E}\{x(n_1)x_2^*(n_2)\}$ that should be $r_x(n_1, n_2) = \mathrm{E}\{x(n_1)x^*(n_2)\}$, since no signal $x_2(n)$ has been defined.

* **Page 114:** In the second entry of Table 3.1 there is an extra left brace in the expression given for Autocorrelation.

* **Page 115:** In the sentence following (3.4.1), the sentence should state "...for all $\zeta \in A$ such that $\Pr\{A\} = 1$, then ...", because the probability of a single outcome of the experiment will, in general, be less than one, $\Pr\{\zeta\} < 1$. This is (correctly) stated in this way in Theorem 3.2.

- **Page 115:** In theorems 3.2–3.4, the book makes the statement that "...the output $y(n, \zeta)$ converges...", but it is not clear what the limit is that results in convergence or what the limits converge to. I believe that the limit that is alluded to, but not stated, is

$$y(n, \zeta) = \lim_{K \to \infty} \sum_{k=-K}^{K} h(k)x(n-k, \zeta). \tag{5}$$

To say that this limit converges is to say that the limit is finite (or has finite variance) for all n as $K \to \infty$. In the limit, we still end up with a random process $y(n, \zeta)$ and we can only talk about what the statistical properties of this sequence converge to as $K \to \infty$. This is discussed in the pages that follow.

* **Page 117:** In calculating the output power just , they apparently used the relation $r_x(-k) = r_x(k)$. This only applies to real signals (see 3.3.29). It's not clear why they made this assumption in this section, but not the sections preceding or following this one. It also seems that with this relation they could have simply left out the two equations that expand and contract $r_h(k)$ and gone directly to (3.4.13) from the first sum.

* **Page 118:** The same error is in (3.4.24). I think this should be

$$\mathrm{E}\left[|y(n)|^2\right] = \sum_{\ell=-\infty}^{\infty} r_x^*(\ell) r_h(\ell) \tag{6}$$

- **Page 119:** The correlation length is poorly described and interpreted. Since the autocorrelation function can be negative, autocorrelations with long memory that oscillate will potentially have much shorter correlation lengths than those that are nonnegative, but brief. I also don't see any justification for the interpretation of correlation length as "...the maximum distance at which two samples are significantly correlated." Clearly $\rho_x(\ell) = 10000.50.5$ for $\ell = 0 - -5$ would not match this definition, since it would have a correlation length of 2.

- **Page 123:** They were a little sloppy with the notation in (3.4.59). Clearly they should have written $r_x(\ell \triangleq j - i)$ as two separate equations: $r_x(j - i) = r_x(\ell)$ where $\ell \triangleq j - i$.

- **Page 124:** They've assumed that $r(\ell)$ has a finite duration in the proof of this theorem. I speculate, but do not know, that this holds in the case that $r(\ell)$ has an infinite duration. One should be cautious, though, that further assumptions may be necessary to ensure convergence and that the linear operations can be reordered.

- **Page 134:** The normalized bias, standard deviation, and mean squared error are generally only meaningful statistics when the parameter θ being estimated is known to only take on non-negative values. For example this may be meaningful for describing the bias and variation of resistance measurements, but it would not be meaniful in describing the bias and variation of bipolar voltages, in most circumstances.

- **Page 134:** Often *mean squared error* is used instead of *mean square error*.

- **Page 135:** The normalized MSE is often defined as

$$\varepsilon \triangleq \frac{\mathrm{MSE}(\theta)}{\operatorname{var} \theta} \tag{7}$$

when the estimate is made simultaneously for multiple values of θ. For example in prediction applications where $\theta = x(n+1)$ $\operatorname{var} \theta$ could be defined as $r_x(0) = \operatorname{var} x(n)$.

However, this makes the most sense when the parameter being estimated is itself a random variable with some distribution.

- **Page 135:** There are variations of the CRLB that do not require the estimator to be unbiased.

- **Page 137:** The book states "We stress that by no means does this imply that a confidence interval includes the true mean with probability of 0.95." This statement is intended to convey the idea that the true parameter θ is not a random variable so we cannot assign a probability to it lying within any region. However, I think the statement could be correctly understood as meaning that the confidence intervals include the mean with a probability of 0.95. Again, the emphasis is that the confidence intervals are the random variables; not the parameter being estimated.

- **Page 138:** In the first paragraph, it is important to note that the random variable $(\hat{\mu}_x - \mu_x)/(\hat{\sigma}_x/\sqrt{N})$ only has a Student's t distribution when the observed signal $x(n)$ is white Gaussian noise. If it is only IID, as stated at the beginning of the section, then the exact distribution of the random variable stated above is unknown. In this case the Student's t distribution is, at best, an approximation.

- **Page 139:** In Example 3.6.1 it should be emphasized that the 95 percent confidence intervals shown in Figure 3.14 are not the confidence interval described in the preceding section, equation (3.6.32). Rather these are confidence intervals based on the Monte Carlo simulation of 10,000 trials and were probably calculated using the knowledge that the estimated mean μ_x is Gaussian an a standard deviation estimated from the 10,000 trials.

- **Page 142:** I do not understand what the dashed lines show in Figure 3.15. I think the only interpretation that makes sense is that these are the *average* confidence intervals averaged over all 10,000 trials and, apparently, centered about the true variance. Keep in mind that the confidence intervals are themselves random variables so you would actually obtain 10,000 different intervals for all 10,000 trials. I think the confidence intervals in this figure are quite misleading since the size of the confidence intervals changes with the true variance and are not constant.

Chapter 4

$*$ **Page 150:** In Figure 4.1 the PSD of the output should be listed as $R_y(\mathrm{e}^{j\omega})$, not $R_x(\mathrm{e}^{j\omega})$. Also the vertical axis of the LTI system's magnitude response should be labelled $|H(\mathrm{e}^{j\omega})|^2$, since it has the same shape as $R_y(\mathrm{e}^{j\omega}) = |H(\mathrm{e}^{j\omega})|^2 R_x(\mathrm{e}^{j\omega})$.

- **Page 150:** The books notes that in the case of a harmonic process the challenging problem is to identify an LTI system that transforms a harmonic process with uniform amplitudes (a flat PSD). I don't think this problem is of much practical interest. For harmonic processes, the problem typically is to determine what the amplitudes, frequencies, and, sometimes, phases are of the different components given a finite segment of a realization that also contains white noise. This problem is sometimes called "frequency estimation."

* **Page 151:** Equation (4.1.8) is wrong. It should be

$$x(n+1) = \underbrace{w(n+1)}_{\text{New Information}} + \underbrace{\sum_{k=-\infty}^{n} h(n+1-k)\left(\sum_{j=0}^{\infty} h_I(j)x(k-j)\right)}_{\text{past information: linear combination of } x(n), x(n-1, \ldots} \tag{8}$$

* **Page 153:** Technically, the sentence following (4.1.20) should say that the "SFM_x is the inverse of the filter *energy*...". The filter does not have a finite *power*.

* **Page 152:** The subscript of the transfer function of the whitening filter in Figure 4.2 should be an "I", not a "1".

* **Page 157:** Equation (4.2.5) holds so long as $n \neq 0$, not just merely $n > 0$. Normally this wouldn't be of any significance since $h(n) = 0$ for $n < 0$, but the text makes a special point that (4.2.5) holds for $n < 0$ at the top of page 159.

- **Page 158:** The statement following (4.2.17) says that they used $r_h^*(-\ell) = r_h(\ell)$, but I don't see how this was used. It seems to me that (4.2.17) follows directly from (4.2.15) with $\ell = 0$.

* **Page 159:** Equation (4.2.21) is wrong. The correct equation is given by (2.3.31) on page 53,

$$R_h(z) = H(z)H^*(z^{-*}) = |d_0|^2 \prod_{k=1}^{P} \frac{1}{(1-p_k z^{-1})(1-p_k^* z)} \tag{9}$$

In the case that $h(n)$ is real, this simplifies to that shown in (4.2.24).

- **Page 159:** I don't understand what the last sentence on this page means. I have no idea what u is (it's not the unit step function) or v are intended to represent in this sentence. If complex conjugate poles are present in the model, they will contribute a two-sided infinite sequence that consists of exponentially damped sinusoids.

- **Page 160:** The notation $\tilde{h}(n)$ is used to denote the response of the system to a periodic impulse train. It might have been more clear to represent the output of the system with $x(n)$ or $y(n)$ in this case, since $h(n)$ is usually reserved solely for the response of the system to an impulse. The motivation for using $\tilde{h}(n)$ was apparently because the input in this case is closely related to an impulse.

- **Page 160:** It may be difficult, at first, to see how (4.2.31) relates to (4.2.30) based on the terse description in the text. Equation (4.2.31) is obtained by removing the first row of (4.2.30) and moving the first column of the remaining matrix to the right side of the equation. Specifically, after removal of the first row we have

$$\begin{bmatrix} r_h(1) & r_h(0) & \cdots & r_h^*(P-1) \\ \vdots & \vdots & \ddots & \vdots \\ r_h(P) & r_h(P-1) & \cdots & r_h(0) \end{bmatrix} \begin{bmatrix} 1 \\ a_1 \\ \vdots \\ a_P \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{10}$$

Once we then move the leftmost column (which is scaled by $a_0 = 1$ in the matrix-vector product) to the right side, we obtain (4.2.31). In written-out form this is given by

$$\begin{bmatrix} r_h(0) & \cdots & r_h^*(P-1) \\ \vdots & \ddots & \vdots \\ r_h(P-1) & \cdots & r_h(0) \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_P \end{bmatrix} = - \begin{bmatrix} r_h(1) \\ \vdots \\ r_h(P) \end{bmatrix}. \tag{11}$$

- **Page 169:** It is interesting to note that the autocorrelation in (4.2.84) is a function of a cosine with a non-zero phase. We know that, for systems with a real impulse response, the autocorrelation is an even function and can be expressed as a sum of cosines. Since (4.2.84) is expressed only for $\ell \geq 0$, this function is still even despite the non-zero phase, but it's not clear how this relates to an expression of $r(\ell)$ as a sum of cosines.

Chapter 5

$*$ **Page 200:** The plots shown in Figure 5.3 are for 8-point and 7-point periodic extensions, but this is inconsistent with the example that refers to them, which calls for 10- and 15-point periodic extensions. However, the point that they illustrate is still conceptually correct.

- **Page 200:** The continuous-time representation of the periodic signal shown in Figure 5.3 (bottom left) is misleading. Here they have joined segments of the periodic continuous-time signal in the same manner that they did so with the discrete-time signal. However, this is not the continuous-time signal that would be obtained if bandlimited interpolation was applied to the discrete-time signal.

- **Page 201:** The idea of sampling in the discrete-time domain discussed on this page and illustrated at the bottom of Figure 5.4 never occurs in practice. This aliasing would only occur if the spectrum was sampled more sparsely over a range of $-\pi < \omega < \pi$ than the number of actual observed samples in the signal segment. Since, in practice, the spectrum is computed by taking the DFT of the signal, the number of samples in the frequency domain is the same as the length of the signal. If zero padding is used, then the number of samples in the frequency domain is greater than the number of samples in the observed segment.

- **Page 203:** The width of the mainlobe is apparently defined as the smallest frequency ω such that $A(\omega) = 0$. That is

$$\omega_{\text{ML}} = \underset{\omega>0}{\operatorname{argmin}} |A(\omega)| \tag{12}$$

- **Page 204:** I don't understand the point of the footnote on this page. If N samples are observed of a signal than most (all?) textbooks would say the signal has a duration of N samples. I don't understand why they list $N-1$ as the "duration" of the data window. Other books use a divisor of N, instead of $N-1$.

- ∗ **Page 205:** The definition of the mainlobe width defined at the bottom of this page as $\triangle\omega = 2\pi/(N-1)$ is inconsistent with the mainlobe width used in (5.1.10) on page 204, which was apparently $2\pi/N$.
- • **Page 205:** Note that idea that the mainlobe width must be smaller than the difference in frequencies of two spectral lines (ω_1 and ω_2) is just a rule of thumb. In practice you could see two distinct peaks with a smaller mainlobe width, but they would be less distinct, similar to the bottom plot in Figure 5.7.
- • **Page 205:** In the first sentence on this page they define an exact value for the mainlobe width $\triangle\omega$, but they have not given a precise definition of how this width is defined, much less calculated. The mainlobe width is actually transition width of a Kaiser window with the same peak approximation error. A more thorough explanation of these values can be found in Chapter 7 of Oppenheim and Schafer's "Discrete-time Signal Processing".
- ∗ **Page 206:** The header for the fourth column of Table 5.1 contains a typo. It says "Exact mainlode width". It should be "Exact mainlobe width".
- ∗ **Page 215:** In the sentence preceding Example 5.3.1 the statement, "…related to the number observations, …" should be "…related to the number of observations, …"
- ∗ **Page 219:** Similarly, in Example 5.3.3 the sentence preceding the second equation should begin, "Since the two frequencies, …"
- ∗ **Page 233:** At the top of the page, the second item in the list of reasons why the periodogram is an unacceptable estimate states, "its variance is equal to the true spectrum $R_x(e^{j\omega})$." This should state that "its variance is proportional to the square of the true spectrum $R_x(e^{j\omega})$."
- ∗ **Page 236:** The text is inconsistent in their spelling of "gage".
- ∗ **Page 243:** The second sentence in Example 5.4.1 ends with "…as shown in Figure 5.27." It should be "…as shown in Figure 5.26."
- • **Page 243:** Equation (5.4.35) is only a coarse approximation of the actual standard deviation of the (PE) estimated coherence and should be used with caution.
- • **Page 243:** The estimated coherence decreases slightly in Fig. 5.27 in the region where the SNR is the highest and the coherence should, theoretically, be maximized. The text suggests that this is due to"bias error due to the lack of sufficient frequency resolution." I question whether this is true.
- • **Page 245-6:** The frequency domain estimates in Figures 5.28 and 5.29 used insufficient zero padding. The linear interpolation used between the coarse sampling in the frequency domain can cause significant errors and should be avoided by using more zero padding.
- • **Page 237-246:** Its disturbing to me that no confidence intervals are included in the practical examples of Sections 5.3 or 5.4. I think this is a critical component of these estimators and good confidence interval estimates are available for almost all of them.

Without them, it is nearly impossible to determine what features of the estimates are statistically significant.

* **Page 245:** The statement, “Thus pressure drops due to a decrease in flow.” is not accurate. The role of the baroreflex response is to sustain pressure, regardless of the vascular resistance or cardiac output. Similarly, the interpretation of the “0.10 Hz” oscillations are controversial and not fully understood. It is not clear if they are solely sympathetic in origin or if they are due to both parasympathetic and sympathetic activity.

Chapter 6

* **Page 281:** The term $r_*(m-k)$ in (6.4.16) should be $r_x(m-k)$.

* **Page 282–285:** There are several problems with Example 6.4.1. First, the frequency ω_0 is assumed to be known and is used in the solution, contrary to the first sentence. Second, because the signal consists of a single tone the optimal filter would have infinite attenuation at all frequencies except ω_0. This problem would be better handled by directly estimating A. An FIR filter, which can only generate $M-1$ zeros, is not well suited to this problem. Third, the statement “We see that an $M = 20$ order filter provides satisfactory performance” would not be true in most practical applications. Clearly as M increases the error will decrease and the magnitude response shown in Figure 6.14 is not a close approximation to the optimal bandpass filter centered at ω_0 with a very narrow passband, which the books calls a “sharp bandpass.” Fourth, f_0 is sometimes used instead of ω_0, without explicitly expressing the relationship between the two frequencies, though most readers would realize $f_0 = 2\pi\omega_0$. Finally, the reader should be aware that the dashed line in Figure 6.15 was just as likely to be lower than the theoretical MMSE shown by the solid line as above it. The point of this figure, as the book correctly states, is merely to show that the filter is optimal across the ensemble, but may not be best for a given realization.

- **Page 297–299:** The spectral factorization in (6.6.15) should probably be $R_x(z) = \sigma_w^2 H_x(z) H_x^*(z^{-*})$ instead of $R_x(z) = \sigma_x^2 H_x(z) H_x^*(z^{-*})$ to make it clear that the σ^2 term is the power of the white noise process driving the putative LTI system $H_x(z)$ and not the power of the process. That is $\sigma_x^2 \neq r_x(0)$ in this case. They used σ_w^2 throughout section 4.1.1 where spectral factorization was first discussed. Note that $r_x(0) \neq \sigma_w^2$, in general.

* **Page 298–299:** The statements after (6.6.22) and (6.6.24) that increasing the order of the filter decreases the MMSE is misleading. These equations only hold for causal and noncausal IIR filters, respectively. If an Mth order FIR filter is used, these equations no longer apply. It is less obvious in the FIR case that the MMSE decreases as M increases, though it still turns out to be true. Note, however, that if the auto- and cross-correlations are estimated, it will turn out that the MMSE can actually increase as M increases due to estimation error.

* **Page 299:** In the last equation of the Solution column of Table 6.5, $h_{\text{nc}}(k)$ should be $h_{\text{c}}(k)$.
- **Page 306:** In the discussion of predictable processes there is an intrinsic assumption that the prediction error filter can only be zero at discrete points. This is certainly true for rational transfer functions which have a finite number of zeros, but it is less obvious in the more general case.
* **Page 308:** The term $\boldsymbol{R}_{y_D x}(\ell) = \ldots$ towards the bottom of the page in Example 6.7.1 should be $\boldsymbol{R}_{y_D x}(z) = \ldots$.
* **Page 310:** In the first paragraph of Section 6.8, "...deviates form the ideal..." should be "...deviates from the ideal..."
* **Page 311:** The second equality in (6.8.2) should be

$$\triangleq \sum_{k=-\infty}^{\infty} a_k \tilde{h}_{\text{r}}(t - kT_B) + \tilde{v}(t) \tag{13}$$

* **Page 320:** Just below (6.9.13) it should state "...for colored additive noise." The same correction should be made in the following paragraph.
* **Page 328:** In problem 6.23 the equation $P^{(d)} = \sigma^2 = \sum_{n=0}^{D-1} |h_x|^2(n)$ should be $P^{(d)} = \sigma^2 = sum_{n=0}^{D-1} |h_x(n)|^2$.

Chapter 7

* **Page 342:** The diagram for the decorrelator in Figure 7.1 appears to be wrong. This does not implement either of the key equations for the whitening equations (7.1.57) or (7.1.53).
* **Page 353:** Equation (7.3.54) should be $P_{\boldsymbol{c}(m+1)} = P_{cm} - \beta^*_{\boldsymbol{c}m} k_{\boldsymbol{c}m}$, not $P_{\boldsymbol{c}(m+1)} = P_{cm} - \beta_{\boldsymbol{c}m} k_{\boldsymbol{c}m}$.
* **Page 355:** I believe the second expression in (7.4.10) should be $= -\boldsymbol{r}_m^{\text{H}} \boldsymbol{J} \boldsymbol{c}_m + d_{m+1}$, not $= -\boldsymbol{c}_m^{\text{H}} \boldsymbol{J} \boldsymbol{r}_m + d_{m+1}$ as given. Of course in the real-valued case, these are the same.
- **Page 379:** I disagree with the statement that the Kalman filter's "...use in statistical signal processing is somewhat limited (adaptive filters ... are more appropriate)." The Kalman filter is widely used in statistical signal processing and adaptive filter algorithms, as discussed in Chapter 10.
* **Page 384:** In Table 7.5, row 3(d), second equation) the term $\boldsymbol{H}^{\text{H}}(n$ is missing the right parenthesis and should be $\boldsymbol{H}^{\text{H}}(n)$.

Chapter 8

- **Page 405–406:** In the deterministic data matrix case they give a very nice, concise proof that an unbiased estimate of the error variance is given by $\frac{1}{N-M} E_{\text{ls}}$ (Page 404, Property 8.2.3). It seems to me that it is critical that they emphasize that this doesn't hold in the stochastic case and explain why. The reader may, in fact, be misled to believe that it does hold in the stochastic case since they end the discussion of this section with the statement, "...the results obtained for the deterministic data matrix $\boldsymbol{X}$ are also valid for the stochastic case." In fact, the only result that is valid is that the least squares estimator of the parameter vector is unbiased in both cases. This may cause significant confusion in Chapter 9 when they apparently use a different (biased) estimate of the error (Page 451, (9.2.13) and (9.2.14)).

∗ **Page 408:** The Prewindowing paragraph should state that it is equivalent to setting $x(-1), \ldots, x(-M+1)$ equal to zero, not $x(0), x(-1), \ldots, x(-M+1)$ equal to zero. The term $x(0)$ is not set equal to zero with this windowing scheme.

Chapter 9

∗ **Page 446:** The second sum in (9.1.1) should have a lower limit of $k = 0$, not $k = 1$ as given.

∗ **Page 448:** The bullet at the bottom of this page suggests that one way of checking the goodness of fit of a model is to make sure the performance criterion (e.g., least squares) decreases fast enough as the model order is increased. In fact, if it decreases too fast it would suggest the model was not a good fit. I think this should be "The criterion of performance does not decrease too fast as we increase the order of the model."

- **Page 450:** In the second sentence following (9.2.9), the sentence begins "Furthermore, if $\text{AR}(P_0)$ is minimum phase ...". However, $\text{AR}(P_0)$ will be minimum phase whenever $\hat{\boldsymbol{R}}$ is positive definite, which it is for every estimator described in the book. Thus, this condition is always satisfied for the estimators described.

∗ **Page 451:** The variance estimates in (9.2.13) and (9.2.14) are inconsistent with the unbiased estimates suggested in Chapter 8 (page 404, (8.2.43). I don't understand why they didn't try to adjust for the lost degrees of freedom.

- **Page 449–462:** They seem to have only partially adopted the notation from Chapter 8 in this section. I don't understand why they switched to specifying the energy of the error with $\mathcal{E}_P$, instead of E_{ls} as they did in Chapter 8.

- **Page 449-462:** They seem to blur the slight distinction between all-pole models (AP) and autoregressive (AR) processes in this chapter and in Figure 9.12 in particular. Instead of consistently referring to the signal models as AR models, they introduce use "all-pole signal model" to mean a white noise signal process with an all-pole system, rather than meaning any signal processed with an all-pole system, as it was defined in Chapter 4 (Page 154).

- **Page 451:** The reader may wonder why they are apparently using biased estimates of the variance of the excitation process in (9.2.13) and (9.2.14), given that an unbiased estimate was obtained in Chapter 8 (Page 404, (8.2.43)). The short answer is that this estimate only holds in the deterministic case when the observations are statistically independent. There is not, to my knowledge, an unbiased estimate of the error in the case of a stochastic data matrix. I think it would be helpful to the reader if this was explained ore fully, perhaps in a footnote. I also think it would be helpful if they stated explicitly that the variance estimates in (9.2.13) and (9.2.14) were negatively biased.

- **Page 451:** In the first section of the chapter they advocate for using ACF and PACS to select the model order but in the first two examples of Section 9.2.1 they merely use it for model verification.

- **Page 453:** The PACS clearly suggested a 5th or even 6th order model for Example 9.2.3, but for unexplained reasons they chose an AR(4) model anyway.

$*$ **Page 454:** Equations (9.2.15) and (9.2.16) are clearly wrong and inconsistent (8.4.13) on Page 414, even when the differences in notation are taken into account. The second product $\bar{\boldsymbol{X}}^{\mathrm{T}}\bar{\boldsymbol{x}}^*$ should be $\boldsymbol{J}\bar{\boldsymbol{X}}^{\mathrm{T}}\bar{\boldsymbol{x}}^*\boldsymbol{J}$.

- **Page 455–457:** It is quite odd to me that they compare the parametric estimates to the Periodogram after stating unequivocally in Chapter 5 that the Periodogram is a poor (the worst?) of the nonparametric estimates. Why didn't they use one of the better nonparametric spectral estimates (e.g., Welch's method or the Blackman-Tukey method) that permits the user to control the bias-variance tradeoff?

- **Page 456:** In Figure 9.10 two of the estimates are labelled "Rectangular window" and "Hamming window". Apparently the "rectangular window" corresponds to the full windowing case, since this type of windowing is listed in the figure caption and example section. It is not at all clear how the Hamming window was applied. Was it a "data window" (Page 203), and "correlation window" (Page 223), or the weights in a weighted least squares estimate (Page 403)? I believe only the last is a reasonable choice because the first two techniques would bias a parametric estimate. However, the description of a "...20 ms, Hamming windowed, speech signal..." for Figure 9.12 would suggest that they actually used a Hamming data window in these examples.

- **Page 456:** As the last of 3 consequences of the spectral representation of the error signal they state that "The all-pole model provides a good estimate of the envelope of the signal spectrum $|X(\mathrm{e}^{j\omega})|^2$." I don't know how they are defining envelope here, but it seems to be based on the example given in Figure 9.12 in which case the AR(28) estimate appears to estimate the envelope of a series of frequency peaks estimated by the Periodogram. I think it is more likely that these are actually harmonics of quasi-periodic components of a speech signal that the AR model is failing to estimate. Clearly this is not a desirable property.