System Identification

Supplementary notes: lecture 13

Roy Smith

13 Subspace Identification

13.1 Deriving the time-aliased impulse response function

We will derive the following,

$$h_k = \frac{1}{N} \sum_{n=0}^{N-1} G(n) e^{j2\pi kn/N}.$$

Substituting the relationship for $G(n)$, which is equal to $G(e^{j\omega_n})$ in the noise-free case,

$$G(n) \; = \; G(e^{j\omega_n}) \; = \; \sum_{i=0}^{\infty} g(i) e^{-j2\pi ni/N},$$

gives,

$$h_k = \frac{1}{N} \sum_{n=0}^{N-1} \sum_{i=0}^{\infty} g(i) e^{j2\pi n(k-i)/N}$$

This is only non-zero for k and i related by,

$$\frac{i2\pi}{N} = \frac{k2\pi}{N} + 2\pi l, \quad l = 0, 1, \ldots,$$

or equivalently, $i = k + Nl$.

So this gives,

$$h_k = \frac{1}{N} \sum_{n=0}^{N-1} \underbrace{\sum_{l=0}^{\infty} g(k+Nl)}_{\text{independent of } n}$$

$$= \sum_{l=0}^{\infty} g(k+Nl)$$

This is a time-aliased version of the impulse response.

This infinite summation can be reformulated by noting that,

$$g(k) = CA^{(k-1)}B,$$

and so,

$$g(k+Nl) = CA^{(k+Nl-1)}B.$$

Summing the impulse response terms gives,

$$\begin{aligned} h_k = \sum_{l=0}^{\infty} g(k+Nl) &= C\left(\sum_{l=0}^{\infty} A^{(k+Nl-1)}\right)B \\ &= CA^{(k-1)}\sum_{l=0}^{\infty} A^{Nl}B. \end{aligned}$$

If A is stable, $\rho(A) < 1$, then,

$$\begin{aligned} h_k &= CA^{(k-1)}\sum_{l=0}^{\infty} A^{Nl}B \\ &= CA^{(k-1)}\left(I - A^N\right)^{-1} B \\ &= CA^{(k-s-1)}\left(I - A^N\right)^{-1} A^s B. \end{aligned}$$

This last relationship holds for any integer s because A clearly commutes with the inverse term. It also shows that H has a Hankel structure.

Furthermore, we can write H directly as,

$$\begin{aligned} H &= \begin{bmatrix} h_1 & h_2 & h_3 & \cdots & h_r \\ h_2 & h_3 & \iddots & \iddots & h_{r+1} \\ h_3 & \iddots & \iddots & \iddots & \\ & \iddots & \iddots & & \vdots \\ h_q & h_{q+1} & & \cdots & h_{q+r-1} \end{bmatrix} \\ &= \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{q-1} \end{bmatrix} \left(I - A^N\right)^{-1} \begin{bmatrix} B & AB & \cdots & A^{r-1}B \end{bmatrix} \end{aligned}$$

This factorisation shows that the range space of H is the extended observability space, $\mathcal{O}$. This is the key feature we require.

13.2 Noise in the irregularly-spaced frequency case

The situation in the presence of noise requires more careful treatment in order to guarantee asymptotic consistency. In the noisy case we would have,

$$\mathcal{G} = \mathcal{O}\mathcal{X}_c + \Gamma\mathcal{W} + \mathcal{V},$$

with the noise matrix $\mathcal{V}$, having the same structure as $\mathcal{G}$. The noise structure can be written as,

$$\mathcal{V} = \mathcal{W}\,\mathrm{diag}\left(V(\mathrm{e}^{j\omega_1}), \cdots, V(\mathrm{e}^{j\omega_N})\right).$$

Now we can express,

$$\begin{aligned}\mathcal{V}_r\mathcal{V}_r^T &= \left[\mathrm{real}(\mathcal{V}) \quad \mathrm{imag}(\mathcal{V})\right]\left[\mathrm{real}(\mathcal{V}) \quad \mathrm{imag}(\mathcal{V})\right]^T \\ &= \mathrm{real}\left(\mathcal{W}\,\mathrm{diag}\left(V(\mathrm{e}^{j\omega_1})V(\mathrm{e}^{j\omega_1})^*, \cdots, V(\mathrm{e}^{j\omega_N})V(\mathrm{e}^{j\omega_N})^*\right)\mathcal{W}^*\right)\end{aligned}$$

This has expected value,

$$E\left\{\mathcal{V}_r\mathcal{V}_r^T\right\} = \mathrm{real}\left(\mathcal{W}\,\mathrm{diag}\left(\phi_v(\omega_1), \cdots, \phi_v(\omega_N)\right)\mathcal{W}^*\right)$$

If we know the noise spectra we can define a weighting,

$$KK^T = \alpha\,\mathrm{real}\left(\mathcal{W}\,\mathrm{diag}\left(\phi_v(\omega_1), \cdots, \phi_v(\omega_N)\right)\mathcal{W}^*\right),$$

for some $\alpha > 0$. The matrix K can be calculated via a Cholesky factorisation.

To achieve asymptotic consistency we solve the weighted problem,

$$K^{-1}\mathcal{G}_r\mathcal{W}_r^{\perp} = K^{-1}\mathcal{O}\mathcal{X}_{cr}\mathcal{W}_r^{\perp} + K^{-1}\mathcal{V}_r\mathcal{W}_r^{\perp}.$$

McKelvey proves that this gives consistency as,

$$K^{-1}\mathcal{V}_r\mathcal{W}_r^{\perp}\left(K^{-1}\mathcal{V}_r\mathcal{W}_r^{\perp}\right)^T \longrightarrow \alpha^{-1}I.$$

Many of the variants of subspace ID (particularly in the time-domain) differ in the treatment of the noise weighting. Note that the results will depend somewhat on being able to estimate the noise spectra. A comprehensive summary can be found in [1].

13.3 Shortcomings of standard SVD-based subspace ID methods

This discussion is taken directly from [2].

There are several difficulties that arise with SVD-based subspace identification.

Determining the appropriate rank of H This is typically done by selecting an index where the singular values drop significantly. Unfortunately it is not always straightforward. The addition of noise to the matrix H makes all of the singular values non-zero and while the noise added to each element of H is relatively small, the cumulative effect on the $n_x + 1$st singular value can be large. This effect is amplified by the Hankel structure of both H and the added noise. Numerical experiments indicate that the variance of the non-zero singular values is higher for matrices with Hankel structure than for unstructured matrices.

The lack of Hankel and shift structure in the truncated SVD Given a Hankel matrix, H, one can find matrices, A, B and C which reconstruct H exactly. In general, if $H \in \mathcal{R}^{n_y N \times n_u N}$, the representation will be of order $2N - 1$. In this case the basis for the range space of H is $U = \begin{bmatrix} U_1 & U_2 \end{bmatrix}$ and U has the required shift structure. However $U_1 \Sigma_1 V_1^T$ is not Hankel and U_1 does not generically have this shift structure. There is no reduced order $\hat{C}$ and $\hat{A}$ from which one can calculate U_1. The least-squares estimate suffers from the problem that errors occur in both the left and righthand factors. Total least-squares gives some improvement but does not account the for the fact that for the majority of matrix components in the equation the errors in the left and righthand sides are actually equal.

The weighting of the effects of the noise The contribution of the noise to the singular values of the Hankel matrix H is not simple. The effect on the individual blocks, $\hat{h}_i$, of H is linear (via the Fourier Transform), but different blocks appear in H a different number of times. One of the consequences of this is that the variance of the effect of noise on the singular values appears to be larger. If one were to weight the effect of individual noise on $\hat{h}_i$, this weighting would have to be applied via a Hadamard multiplication. Multiplicatively weighted SVD problems will not correctly account for such noise.

References

[1] L. Ljung, *System Identification: Theory for the User*, 2nd ed. Prentice-Hall, 1999.

[2] R. S. Smith, "Frequency Domain Subspace Identification Using Nuclear Norm Minimization and Hankel Matrix Realizations," *IEEE Transactions on Automatic Control*, vol. 59, no. 99, pp. 2886–2896, 2014.