

Some Results on Identifying Linear Systems Using Frequency Domain Data

Lennart Ljung
Department of Electrical Engineering
Linköping University
S-581 83 Linköping, Sweden

Abstract

The usefulness of frequency domain interpretations in linear systems is well known. In this contribution the connections between frequency domain and time domain expressions will be discussed. In particular, we consider some aspects of using frequency domain data as primary observations.

1 Introduction

For linear systems the connections and interplay between time-domain and frequency domain aspects have proved to be most fruitful in all applications. We shall in this contribution discuss some aspects in applications to linear system identification.

There are two sides of this interplay. One is to consider the primary observation to be in the time-domain, and then to interpret corresponding identification criteria, algorithms and properties in the frequency domain. There are many early results of this character, e.g. [9], [2], [1], [4]. More recently such results have been exploited and developed in [6].

The other side of the interplay is to consider the primary observations to be in the frequency domain. That is, the Fourier transforms of the measured signals (or certain ratios of them) are treated as the actual measurements. This view has been less common in the traditional system identification literature, but has been of great importance in the Mechanical Engineering community, vibrational analysis and so on. An early reference is [5]. An excellent recent account, with many references, of this view is given in the book [7].

This contribution will deal with a few questions of the latter view from a more traditional System Identification background.

2 Parameterized models

We shall throughout this paper consider linear models in discrete or continuous time, parameterized as follows:

$$y(t) = G(q, \theta)u(t) + H(q, \theta)e(t) \quad (1)$$

(Discrete time)

$$y(t) = G(p, \theta)u(t) + H(p, \theta)e(t) \quad (2)$$

(Continuous time)

Here y , u and e are the output, the input and the noise source, respectively. e is supposed to be white noise with variance (intensity) λ . q is the shift operator and p is the differentiation operator.

A typical parameterization, both in continuous and discrete time could be as a rational function

$$G(p, \theta) = \frac{b_1 p^{n-1} + \dots + b_n}{p^n + f_1 p^{n-1} + \dots + f_n} = \frac{B(p)}{F(p)} \quad (3)$$

$$\theta = (b_1, \dots, b_n, f_1, \dots, f_n).$$

In discrete time we could, e.g. also use parameterizations that originate from an underlying continuous time state space model, discretized under the assumption that the input is piecewise constant over the sampling interval:

$$G(q, \theta) = C(qI - e^{A(\theta)T})^{-1} \int_0^T e^{A(\theta)\tau} B(\theta) d\tau \quad (4)$$

See, e.g. [6] for many more examples of the parameterization (1).

3 Time domain data

Suppose input-output data in the time domain are given:

$$z^N = \{y(t), u(t); t = T, 2T, \dots, NT\} \quad (5)$$

form the corresponding predictions

$$\hat{y}(t|\theta) = G(q, \theta)u(t) + (I - H^{-1}(q, \theta))(y(t) - G(q, \theta)u(t)) \quad (6)$$

and the associated prediction errors:

$$\varepsilon(t, \theta) = y(t) - \hat{y}(t|\theta) = H^{-1}(q, \theta)(y(t) - G(q, \theta)u(t)) \quad (7)$$

and then compute

$$\hat{\theta}_N = \arg \min_{\theta} \sum_{t=1}^N \varepsilon^2(t, \theta) \quad (8)$$

Most frequency domain interpretations of this time domain method go back to the application of Parseval's relationship to the right hand side of (8):

$$\hat{\theta}_N \sim \arg \min_{\theta} \int_{-\pi}^{\pi} |E(\omega, \theta)|^2 d\omega \quad (9)$$

where E is the Fourier transform of ε :

$$E(\omega, \theta) = H^{-1}(e^{i\omega}, \theta)[Y(\omega) - G(e^{i\omega}, \theta)U(\omega)] \quad (10)$$

$$Y(\omega) = \frac{1}{\sqrt{N}} \sum_{t=1}^N y(t)e^{-i\omega t} \quad (11)$$

and similarly for $U(\omega)$. If we introduce the "Empirical Transfer Function Estimate",

$$\hat{G}(e^{i\omega}) = \frac{Y(\omega)}{U(\omega)} \quad (12)$$

(8) - (11) can be rewritten

$$\hat{\theta}_N \sim \arg \min_{\theta} \int_{-\pi}^{\pi} |\hat{G}(e^{i\omega}) - G(e^{i\omega}, \theta)|^2 \frac{|U(\omega)|^2}{|H(e^{i\omega}, \theta)|^2} d\omega \quad (13)$$

4 Frequency domain data

Suppose now that the original data are supposed to be

$$Z^N = \{Y(\omega_k), U(\omega_k), k = 1, \dots, N\} \quad (14)$$

where $Y(\omega_k)$ and $U(\omega_k)$ either are the discrete Fourier transforms of $y(t)$ and $u(t)$ as in (11) or are considered as Fourier transforms of the underlying continuous signals:

$$Y(\omega) = \int_{-\infty}^{\infty} y(t)e^{-i\omega t} dt \quad (15)$$

(or a normalized version). Which interpretation is more suitable depends of course of the signal character, sampling interval and so on.

(13) it would be tempting to use

$$\hat{\theta}_N = \arg \min_{\theta} V(\theta)$$

$$V(\theta) = \sum_{k=1}^N |Y(\omega_k) - G(e^{i\omega_k T}, \theta)U(\omega_k)|^2 \cdot \frac{1}{|H(e^{i\omega_k T}, \theta)|^2} \quad (16)$$

(replacing $e^{i\omega_k T}$ by $i\omega_k$ for the continuous-time model (2).)

If H in fact does not depend on θ (*fixed* or *known noise model*) experience shows that (16) works well. Otherwise the estimate $\hat{\theta}_N$ may not be consistent.

To find a better estimator we turn to the maximum likelihood (ML) method for advice: (We give the expressions for the continuous time case; in the case of (1), just replace $i\omega_k$ by $e^{i\omega_k T}$)

If the data were generated by

$$y(t) = G(p, \theta)u(t) + H(p, \theta)e(t)$$

the Fourier transforms would be related by

$$Y(\omega) = G(i\omega, \theta)U(\omega) + H(i\omega, \theta)E(\omega) \quad (17)$$

To be true, (17) should in many cases contain an error term that accounts for the fact that the measured data $Y(\omega_k)$ often are not exact realizations of (15). For periodic signals, observed over an integer number of periods, (17) may however hold exactly.

Now, if $e(t)$ is white noise, its Fourier transform (suitably normalized) will have a (complex) Normal distribution:

$$E(\omega) \in N(0, \lambda I) \text{ complex} \quad (18)$$

This means that the real and imaginary parts are each normally distributed, with zero means and variances λ . The real and imaginary parts are independent and, moreover, $E(\omega_1)$ and $E(\omega_2)$ are independent for $\omega_1 \neq \omega_2$. This implies that

$$Y(\omega_k) \in N(G(i\omega_k, \theta)U(\omega_k), \lambda |H(i\omega_k, \theta)|^2) \quad (19)$$

according to the model, so that the negative logarithm of the likelihood function becomes

$$V_N(\theta) = \sum_{k=1}^N \left\{ 2 \log |H(i\omega_k, \theta)| + \frac{1}{\lambda} |Y(\omega_k) - G(i\omega_k, \theta)U(\omega_k)|^2 \right\} + N \log \lambda \quad (20)$$

The ML estimate is

$$\hat{\theta}_N = \arg \min_{\theta} V_N(\theta) \quad (21)$$

obtain

$$\hat{\theta}_N = \arg \min_{\theta} \left[N \cdot \log W_N(\theta) + 2 \sum_{k=1}^N \log |H(i\omega_k, \theta)| \right] \quad (22)$$

$$W_N(\theta) = \frac{1}{N} \sum_{k=1}^N |Y(\omega_k) - G(i\omega_k)U(\omega_k)|^2 \cdot \frac{1}{|H(i\omega_k, \theta)|^2} \quad (23)$$

$$\hat{\lambda}_N = W_N(\hat{\theta}_N) \quad (24)$$

Compared to (16) we thus have an additional term

$$\sum_{k=1}^N \log |H(i\omega_k, \theta)|^2 \quad (25)$$

We may note that for any monic, stable and inversely stable transfer function $H(q, \theta)$ we have

$$\int_{-\pi}^{\pi} \log |H(e^{i\omega}, \theta)|^2 d\omega \equiv 0 \quad (26)$$

This is the reason why (25) is missing from criteria that use dense, equally spaced frequencies ω_k for discrete time models (like (13)).

[In fact (25) is the determinant from the change of variables from Y to E (outputs to innovations). In the discrete time domain this transformation is a triangular operator with 1's along the diagonal ($e(t) = y(t)$ -past data). Hence this transformation has a determinant equal to 1, so it does not affect the ML criterion.]

It is apparently often assumed (as in [7]) that the noise model is given or known. Then of course the term (25) is again not essential.

5 Asymptotic properties

The asymptotic properties (as $N \rightarrow \infty$) of the estimate (20)-(21) can be developed in a rather straightforward fashion, using the standard techniques. We confine ourselves below to the case of a fixed noise model $H(i\omega, \theta) = H_*(i\omega)$ and a known λ . Suppose, as $N \rightarrow \infty$ the frequencies ω_k cover the frequency interval $[-\Omega, \Omega]$ with a density function $W(\omega)$. [That is, let $w_N(\Omega_1, \Omega_2)$ be the number of observed frequencies in the interval Ω_1 to Ω_2 when the total number of frequencies is N . Then

$$\lim_{N \rightarrow \infty} \frac{1}{N} w_N(\Omega_1, \Omega_2) = \int_{\Omega_1}^{\Omega_2} W(\omega) d\omega.]$$

formly in θ and with probability 1 to

$$\bar{V}(\theta) = \int_{-\Omega}^{\Omega} |G_0(i\omega) - G(i\omega, \theta)|^2 \frac{\Phi_u(\omega)W(\omega)}{|H_*(i\omega)|^2} d\omega \quad (27)$$

where G_0 is the true transfer function, and $\Phi_u(\omega)$ is the input spectrum. Hence

$$\hat{\theta}_N \rightarrow \arg \min_{\theta} \bar{V}(\theta) \text{ w.p.1 as } N \rightarrow \infty \quad (28)$$

If there exists a value θ_0 such that $G_0(i\omega) = G(i\omega, \theta_0)$ and $\Phi_u(\omega)W(\omega)$ is different from zero at sufficiently many frequencies it will follow that

$$\hat{\theta}_N \rightarrow \theta_0 \text{ as } N \rightarrow \infty$$

In that case the covariance matrix of $\hat{\theta}_N$ will be, asymptotically,

$$Cov \hat{\theta}_N \sim \lambda \cdot \left[\sum_{k=1}^N \frac{G'_{\theta}(i\omega_k, \theta_0)G'_{\theta}(i\omega_k, \theta_0)^* \Phi_u(\omega_k)}{|H_*(i\omega_k)|^2} \right]^{-1} \quad (29)$$

Here G'_{θ} is the gradient of $G(i\omega, \theta)$ with respect to θ and superscript $*$ denotes complex conjugation and matrix transpose.

6 Some practical aspects

There are several distinct features with the direct frequency domain approach that could be quite useful. We shall list a few (see also, e.g., [7])

- **Prefiltering** is known as quite useful in the time-domain approach. For frequency domain data it becomes very simple: It just corresponds to assigning different weights to different frequencies, which in turn is the same as using a frequency dependent $\lambda \equiv$ cheating on the assumed noise levels. It is of course particularly easy to implement perfect band-pass filtering effects in the frequency domain approach.
- **Condensing large data sets.** When dealing with systems with a fairly wide spread of time constants, large data sets have to be collected in the time domain. When converted to the frequency domain they can easily be condensed, so that, for example, logarithmically spaced frequencies are obtained. At higher frequencies one would thus decimate the data, which involves averaging over neighbouring frequencies. Then the noise level (λ_k) is reduced accordingly.
- **Combining experiments.** Nothing in the approach of Section 4 says that the frequency response data at different frequencies have to come

quencies involved ($\omega_k, k = 1 \dots N$) all have to be different. It is thus very easy to combine data from different experiments.

- **Periodic inputs.** The main drawback with the frequency domain approach is that the underlying frequency domain model (17) is strictly correct only for a periodic input and assuming all transients have died out. On the other hand, typical use of the time domain method (6) - (8) assumes inputs and outputs prior to time 0 to be zero. Whichever assumption about past behaviour is closer to the truth should thus affect the choice of approach.
- **Band-limited signals.** If the actual input signals are band-limited, (like no power above the Nyquist frequency) the continuous time Fourier transform (15) can be well computed from sampled data. It is then possible to directly build continuous-time models without any extra work.
- **Continuous-time models.** The comment above shows that direct continuous-time system identification from "continuous-time data" can be dealt with in a much more relaxed way than in the time-domain, with all its mathematical intricacies.
- **Trade-off noise/frequency resolution.** The approach also allows for a more direct and frequency dependent trade-off between frequency resolution and noise levels. That will be done as the original Fourier transform data are decimated to the selected range of frequencies $\omega_k, k = 1, \dots, N$.

7 Some algorithmic questions

The criterion (22) to be minimized is non-quadratic in θ in most cases. This calls for iterative search procedures for the calculation of $\hat{\theta}_N$. This in turn raises two questions:

1. What method should be applied for the iterations?
2. At what parameter values should the search be initialized?

We shall deal with these questions in order.

Iterative minimization

If the noise model H is fixed (θ -independent), the remaining criterion to be minimized in $W_N(\theta)$, which is

deal with such a function minimization is the damped Gauss-Newton method [3]. This apparently is still the best approach around, and is the basic method used in System Identification. Indeed, the MATLAB Signal Processing Toolbox commands for solving (22) for a fixed noise model (`invfreqz` and `invfreqs`) implement this approach.

Unfortunately, it turns out that the additional term (25) may seriously deteriorate the performance of the damped Gauss-Newton procedure. This is, not unexpectedly, most pronounced for continuous time models and for very unequally spaced frequency samples. One probably then has to go to full Newton-methods, which however puts greater demands on the line search. Also, it is important to scale the parameterization, so that the criterion remains reasonably well conditioned.

Initial parameter estimates

Also in the time-domain approach it is very important to provide the Gauss-Newton iterative scheme with good initial conditions. In [6] (Section 10.5) several steps to achieve such initial estimations are described. They are based on the Instrumental Variable (IV) method and the so called repeated Least Squares (rLS) method (i.e. estimating a high order ARX-model, then compute the innovations from this and use them as measured inputs in the next step).

Fortunately these methods can be more or less directly carried over to direct frequency domain methods. *The IV method* (see also [8]) can be described as follows: The problem is to find an initial estimate

$$\hat{G}^{(0)}(e^{i\omega}) = \frac{B(e^{i\omega})}{A(e^{i\omega})}$$

Step i): Solve

$$\min_{a_i, b_i} \sum_k |A(e^{i\omega_k})Y(\omega_k) - B(e^{i\omega_k})U(\omega_k)|^2$$

for A^s, B^s . Let $\hat{G}^s = \frac{B^s}{A^s}$

Step ii): Solve

$$0 = \sum_k (A(e^{i\omega_k})Y(\omega_k) - B(e^{i\omega_k})U(\omega_k)) \cdot \zeta(\omega_k) \quad (30)$$

for A and B where

$$\zeta(\omega_k) = \begin{bmatrix} \vdots \\ \hat{G}^s(e^{i\omega_k}) \cdot e^{il\omega_k} U(\omega_k) \\ \vdots \\ e^{il\omega_k} U(\omega_k) \\ \vdots \end{bmatrix} \quad (31)$$

tively. The vector (31) is the vector of instruments.

The rLS method. is as follows in the frequency domain. The problem is to find $A(q)$ and $C(q)$ in an ARMA model

$$A(q)y(t) = C(q)e(t).$$

Step 1). Solve

$$\min_{\alpha} \sum_R |\alpha(e^{i\omega_k})Y(\omega_k)|^2$$

for $\hat{\alpha}(e^{i\omega})$ for a "high order" polynomial α .

Step 2). Treat

$$\hat{E}(\omega_k) = \hat{\alpha}(e^{i\omega_k})Y(\omega_k)$$

as measured input and solve

$$\min_{A,C} \sum_k |A(e^{i\omega_k})Y(\omega_k) - (C(e^{i\omega_k}) - 1)\hat{E}(\omega_k)|^2 \quad (32)$$

for \hat{A} , \hat{C} . It is my experience that these start-up procedures work well.

8 Conclusions

We have in this contribution discussed various aspects of frequency domain methods for linear system identification. Generally speaking, it could be said that the direct frequency domain approach has been underutilized in conventional system identification. The contribution has been partly of tutorial character, summarizing some main points. In addition the author's experiences with various implementations of the algorithms have been described.

References

- [1] H. Akaike. Maximum likelihood identification of gaussian autoregressive moving average models. *Biometrika*, 1973.
- [2] E. J. Hannan. *Multiple Time Series*. Wiley, New York, 1970.
- [3] J.E.Dennis and R. B. Schnabel. *Numerical methods for unconstrained optimization and nonlinear equations*. Prentice-Hall, 1983.
- [4] P. V. Kabaila and G. C. Goodwin. On the estimation of the parameters of an optimal interpolator when the class of interpolators is restricted. *SIAM J. Control and Optimization*, 1980.

Automatic Control, 1959.

- [6] L. Ljung. *System Identification - Theory for the User*. Prentice-Hall, Englewood Cliffs, N.J., 1987.
- [7] J. Schoukens and R. Pintelon. *Identification of Linear Systems: A Practical Guideline to Accurate Modeling*. Pergamon Press, London (U.K.), 1991.
- [8] A. van den Bos. Identification of continuous-time systems using multiharmonic test signals. *Identification of Continuous-Time Systems*, 1992. Edited by Sinha and Rao, Kluwer Academic, Dordrecht (The Netherlands).
- [9] P. Whittle. *Hypothesis Testing in Time Series Analysis*. PhD thesis, Uppsala University, 1951.

/ingegerd/lennart/cdc93.tex