Pergamon

PII: S0038–092X(01)00052–4

Solar Energy Vol. 71, No. 4, pp. 225–231, 2001

0038-092X/01/$ - see front matter

www.elsevier.com/locate/solener

MEAN PRESERVING ALGORITHM FOR SMOOTHLY INTERPOLATING AVERAGED DATA

M. D. RYMES and D. R. MYERS†
National Renewable Energy Laboratory, 1617 Cole Blvd., Golden, CO 80401, USA

Received 17 April 2000; revised version accepted 11 April 2001

Communicated by RICHARD PEREZ

Abstract—Hourly mean or monthly mean values of measured solar radiation are typical vehicles for summarized solar radiation and meteorological data. Often, solar-based renewable energy system designers, researchers, and engineers prefer to work with more highly time resolved data, such as detailed diurnal profiles, or mean daily values. The object of this paper is to present a simple method for smoothly interpolating averaged (coarsely resolved) data into data with a finer resolution, while preserving the deterministic mean of the data. The technique preserves the proper component relationship between direct, diffuse, and global solar radiation (when values for at least two of the components are available), as well as the deterministic mean of the coarsely resolved data. Examples based on measured data from several sources and examples of the applicability of this mean preserving smooth interpolator to other averaged data, such as weather data, are presented.

1. INTRODUCTION

Applications of solar energy to renewable energy technologies such as photovoltaics, solar thermal systems, daylighting, and passive solar design require site-specific solar radiation resource data and concomitant meteorological data for design and performance evaluation. Analysis of such renewable energy systems for different time scales ranging from annual to minute-by-minute time series are often needed. Obtaining high resolution resource and environmental data to perform system simulations is sometimes difficult. This paper describes a method for interpolating averaged data to any desired resolution, using simple arithmetic.

2. THE PROBLEM

Highly summarized solar radiation and meteorological data is generally readily available. Examples include annual and monthly means of meteorological or solar data, or monthly mean daily total solar resource data. A large amount of important information is lost in the aggregation of data into statistical summaries. No method, including ours, can purport to recreate the individual data points used to generate the aggregated data. Our goal is to produce interpolated data that preserves averaged data values when integrated, is smooth, and obeys user set amplitude constraints (if needed).

Many models generate artificial time series mimicking the natural variability of seasonally fluctuating solar data and meteorological data. Examples include autoregression models (Mora-Lopez and Sidrach-de-Cardona, 1998), models involving cumulative frequency distributions (Gelegenis, 1999), and harmonic analysis (Epstein, 1991; Bouhaddou *et al*., 1997; Knight *et al*., 1990).

Approaches such as linear interpolation between average values can produce over- and under-estimates at subintervals for highly variable data, and the average of the interpolated data will not generally equal the original averaged data. Fig. 1 compares hourly means of 1-min linearly interpolated direct solar radiation data with the original hourly mean data. The figure also shows the absolute error in W/m^2 in the hourly mean of the interpolated data with respect to the original data (gray line). Errors approach 100 W/m^2 in this case.

Cubic spline interpolation leads to difficulties because of implicit assumptions about the 'smooth' nature or lack of discontinuities in the *derivatives, or slopes* of the underlying data (Diggle, 1990). These methods were developed to address interpolation of tabulated analytic functions (Hamming, 1986).

†Author to whom correspondence should be addressed. Fax: +1-303-384-6768; e-mail: daryl_myers@nrel.gov

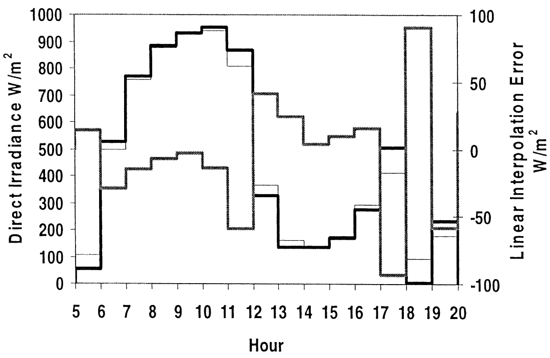

Fig. 1. Hourly mean direct solar radiation data (thick black line) is interpolated to 1-min resolution, and the hourly mean of the interpolated data is shown (thin black line). Gray line is absolute error between hourly means of original and interpolated data hourly means (right axis).

While the above mentioned autoregression methods can be applied to determining many of the characteristic *statistical* properties of high resolution measured data from averaged data, assembling reasonable higher resolution time series sequences of data from aggregated data that preserve statistical properties can be time consuming and complicated. Solar radiation averages may be used to 'scale' extraterrestrial solar radiation on a sub-average interval (by trial and error) so as to retain the proper average; however such scaling parameters do not necessarily exist for meteorological (or other) variables of interest.

Our method resolves a vector of such averaged data, AVG(k); ($k = 1$ to K), into a vector of $K \times N$ sub elements $M(i)$ ($i = 1$ to $K \times N$) which is smooth, having no discontinuities in the slope of the data.

After many frustrations with the wild excursions that spline fitting algorithms can result in, one of the authors (Rymes) developed this 'brute force' method. The advantages of the method are: (1) simple, robust algorithm needing only arithmetic operations; (2) the resulting curve is forced to be smooth (i.e. no discontinuities in slope); (3) the resulting curve is forced to integrate to the original supplied values; (4) user supplied upper and lower limits of any kind, i.e. 'do not exceed average$\times$1.25'; as well as fixed limits (e.g. 'greater than 0, less than 1300') are honored; and (5) the degree of smoothness is selectable.

3. THE METHOD

The method is a simplification of the Markov 'moving average' autoregression approach. We refer to the interpolated data derived from averaged data as 'smoothed' data, and the interpolation itself as 'smoothing'.

The algorithm is recursive in N, the desired resultant data set size. For example, if AVG[24] is the input array of 24 hourly averages, and 1-min data that retain the hourly averages is desired, $M[1440]$ is the output array of $N = 1440$ 1-min values.

For each desired subinterval (minute) we compute the average of each point and the two adjacent points:

$$MN[i] = (MO[i-1] + MO[i] + MO[i+1])/3 \tag{1}$$

MO is updated by replacing the values of MO with MN at each sub-interval. If a day's worth of data is considered in isolation, the endpoint values wrap around midnight:

$$MN[0] = (MO[1439] + MO[0] + MO[1])/3$$

and

$$MN[1439] = (MO[1438] + MO[1439] + MO[0])/3. \tag{2}$$

If formulated in a matrix equation

$$A \cdot MO = MN \tag{3}$$

A is a tri-diagonal matrix with the diagonal and first off-diagonal elements equal to $1/3$. The matrix A is Markovian and because the diagonal is non-zero, *convergence to a stable solution is guaranteed* (Noble and Daniel, 1988).

The suggested number of iterations, N, is the number of output data values. After the Nth iteration, each data value contains data trinomially averaged from all the original data.

4. UNBOUNDED SMOOTHING

Unbounded smoothing implies no concern about large excursions or physical limits. We distribute corrections for the difference between initial averages and means of interpolated data uniformly over all values composing the average.

We add a correction term $C(K)$ to each $MN(i)$, where $C(K)$ is the average difference between initial average and new minute values:

$$C(K) = \sum[\mathrm{AVG}(K) - MN(i)]/N. \tag{4}$$

The sum is over all i within averaging interval K.

$C(K)$ is used to make the average of the corrected $M(i)$ equal to each initial AVG(K) by replacing each $MN(i)$ with $MN(i) - C(K)$ within

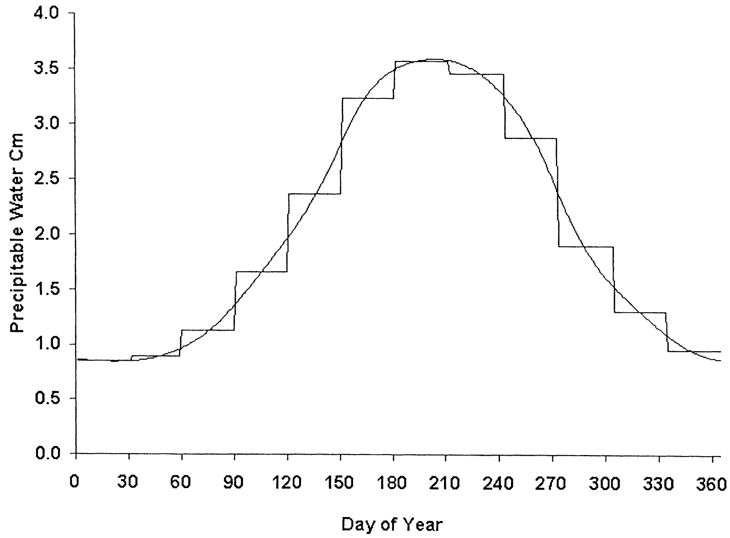

Fig. 2. Monthly averaged precipitable water vapor (step curve) smoothly interpolated to daily values.

each K average interval. The index i is incremented and we repeat the trinomial averages according to Eqs. (1) and (2). This completes the smooth interpolation of the data.

5. BOUNDED SMOOTHING

Limiting the smoothed data to prescribed limits (fixed or variable) is often desirable. A simple approach is to replace smoothed values exceeding the given limits with the limit itself. However this can lead to discontinuities (sharp corners).

Our solution was to implement a uniform distribution of correction terms over every sub-element in a manner similar to the unconstrained case. To constrain each $MN(i)$ to be less than a maximum value MAX, set $MN(i) = \mathrm{MAX}$ for all i where $MN(i) > \mathrm{MAX}$.

If the smoothed values $MN(i)$ require upward correction, make the correction $F(K)$ a fixed fraction of the difference $[\mathrm{MAX} - MN(i)]$ with respect to the initial difference $[\mathrm{MAX} - \mathrm{AVG}(K)]$ using

$$F(K) = \sum[\mathrm{MAX} - \mathrm{AVG}(K)] / \sum[\mathrm{MAX} - MN(i)] \tag{5}$$

and correct the sub-element:

$$MN[i] = \mathrm{MAX} - F(K) \times (\mathrm{MAX} - MN[i]). \tag{6}$$

Similarly, for a lower bound MIN, set $MN(i) = \mathrm{MIN}$ for all i where $MN(i) < \mathrm{MIN}$. If the sum of

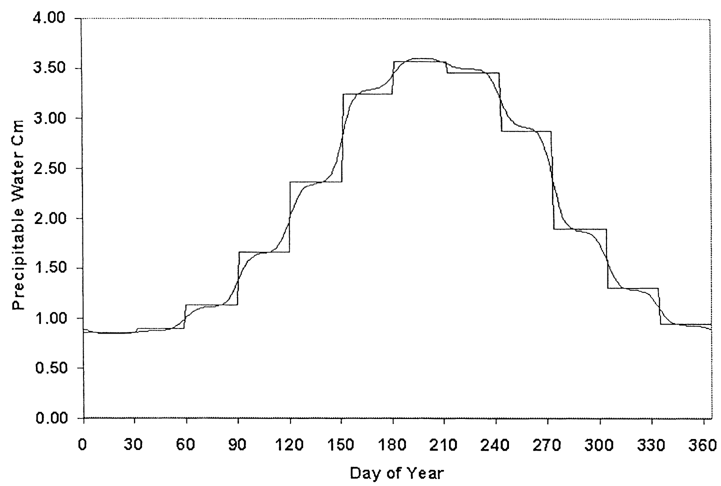

Fig. 3. Same as Fig. 2, but with only eight rather than 365 iterations of smoothing algorithm applied.

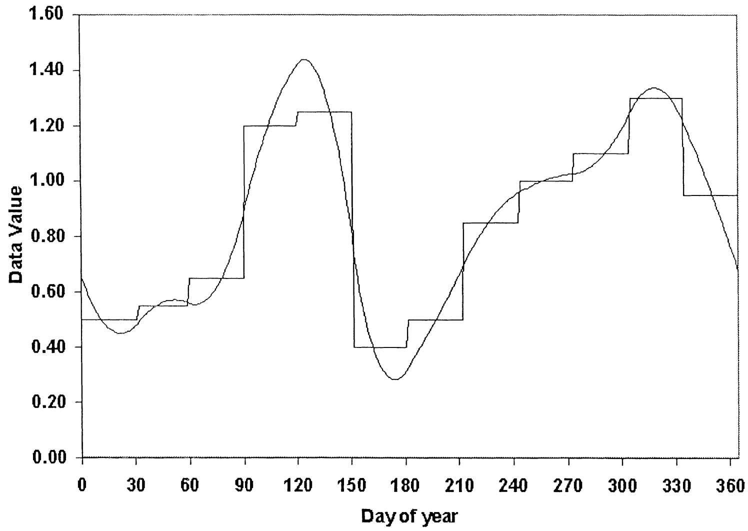

Fig. 4. Unconstrained smoothing (smooth line) of averaged data with large step amplitudes.

the corrected $MN(i)$ is too high with respect to initial AVG(K) $MN(i)$, and require downward correction,

$$F(K) = \sum[MN(i) - \mathrm{AVG}(K)] / \sum[MN(i) - \mathrm{MIN}] \quad (7)$$

and

$$MN[i] = \mathrm{MIN} + F(K) \times (MN[i] - \mathrm{MIN}). \quad (8)$$

All sums are over all i within averaging interval K.

6. EXAMPLES

Fig. 2 shows results of an unconstrained daily smoothing of monthly precipitable water data from Washington, DC given by Gueymard (1994). The monthly values (contained in array AVG[12]) are represented by the step function and the smoothed daily values are represented by the smooth curve.

A smaller number of iterations bevels the edges of the initial step-function. For example, in Fig. 2, 365 iterations produced the smooth curve. In Fig. 3, eight iterations produced the rounded curve.

An example of bounded smoothing for an

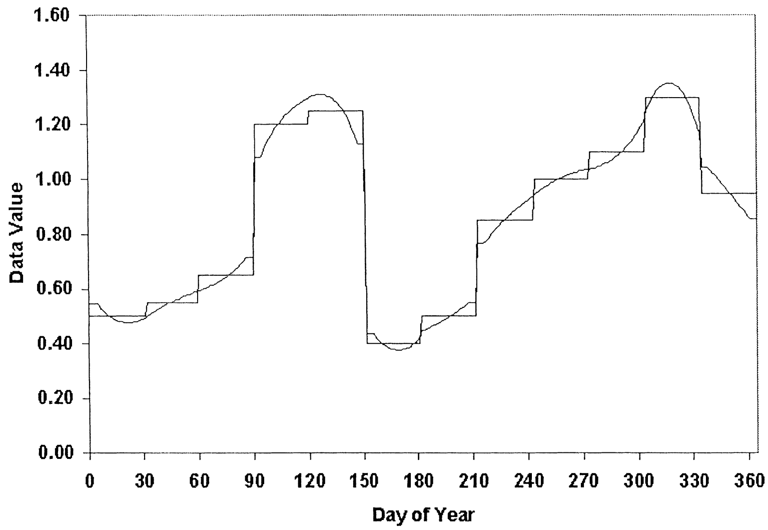

Fig. 5. Average data of Fig. 4 smoothly interpolated using upper and lower bound constraints.

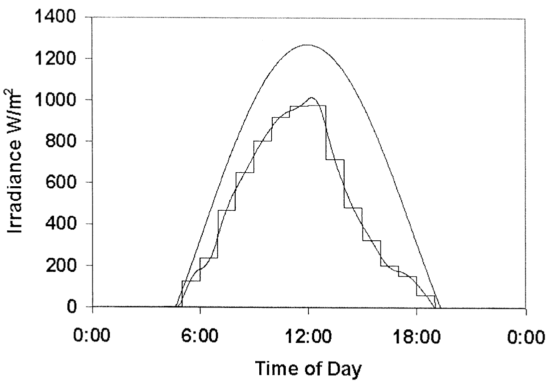

Fig. 6. NREL SRRL global horizontal hourly average data smoothly interpolated to 1-min resolution using bounds of extraterrestrial global horizontal irradiance and zero.

artificial data set of monthly means, interpolated to daily values, is shown in Figs. 4 and 5.

In Fig. 4, unbounded smoothing overshoots the highest and lowest average values. In Fig. 5, bounds were to within 10% of the monthly average for all days d in corresponding months m. That is, $\mathrm{MAX}[d]=1.1\cdot\mathrm{AVG}[m]$ and $\mathrm{MIN}[d]=0.9\cdot\mathrm{AVG}[m]$.

7. SMOOTHING THREE-COMPONENT HOURLY BROADBAND SOLAR IRRADIANCE DATA

When individual time series data are linked, care must be taken that the linkage be preserved. Total global horizontal (T), direct normal beam (B), and diffuse horizontal (D) solar irradiance data are linked through the following equation:

$$T = D + B \cdot \cos(Z), \tag{9}$$

where Z is the solar zenith angle. An independent smoothing of T, B, and D does not ensure that the coupling equation will be valid. Because D and B are components of T, global horizontal irradiance should be smoothed first.

In the examples below, bounds were set to the extraterrestrial (top of atmosphere) global horizontal irradiance for MAX, and zero for MIN. Figs. 6 and 7 show 1-min smoothing of the June 1, 1999 hourly average global horizontal and direct normal irradiance data from NREL's Solar Radiation Research Laboratory (SRRL). The upper curve of Fig. 6 is MAX[m], corresponding to the extraterrestrial global horizontal irradiance.

Direct normal irradiance in Fig. 6 is modeled using the smoothed T curve as an upper bound. Because B differs qualitatively by the cosine of the solar zenith angle, each 1-min value of MAX[m] is set to (smoothed-$T[m]/\cos Z[m]$), or a 'normal' variant of the global horizontal. Again, MIN is zero.

This example shows how flexible upper bounds may be. In fact, this specific boundary may be risky at high zenith angles, due to the small values in the denominator. It is possible that unrealistic large direct beam values may result. Note also that the MAX restriction requires the global and direct values to be *properly measured*; an hourly value of direct irradiance on the horizontal that exceeds the total will create convergence problems for the smoother.

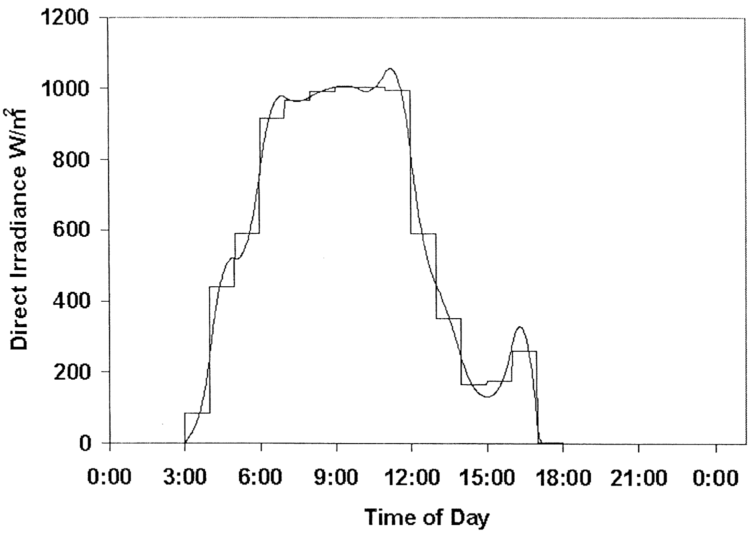

Fig. 7. NREL SRRL hourly average direct normal irradiance smoothly interpolated using the smoothed total curve of Fig. 6 divided by cos(Z) as an upper bound.

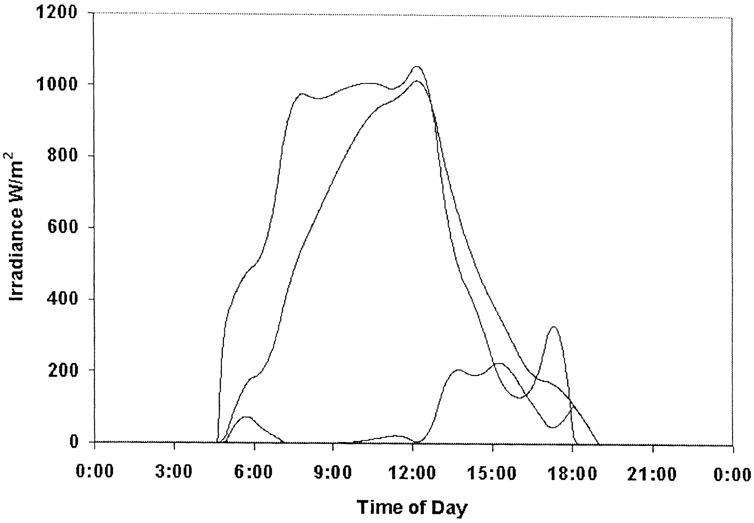

Fig. 8. Global horizontal, direct, and diffuse smoothly interpolated 1-min averages for hourly averaged SRRL data for 1 June 1999.

Finally, diffuse horizontal is calculated from the linking Eq. (9). Because direct normal irradiance has been smoothed with no assurance that the hourly direct horizontal irradiance values have been preserved, 1-min diffuse horizontal irradiance values may not match their hourly averages.

Minor inconsistencies in this regard are preferable to independently smoothed irradiance components that do not correlate with each other. Fig. 8 displays the estimates of 1-min global horizontal, direct normal, and diffuse horizontal solar irradiance at SRRL on June 1, 1999. The maximum difference between the original diffuse irradiance and the diffuse data averaged from the smooth 1-min D data is 50 W/m^2, or about 10%. This is because individual instrument characteristics are not accounted for. The example serves to warn that coupled smoothing of qualitatively different data will produce errors.

8. SUMMARY

We described a simple arithmetic algorithm for producing smooth higher resolution data from averaged data. The method does not recreate the exact underlying data used to construct the average data. This is illustrated by comparing Fig. 9, the actual 1-min direct normal irradiance data from SRRL on June 1, 1999, with Figs. 7 and 8. We remove artificially abrupt transitions in input data, enabling researchers to focus on the features

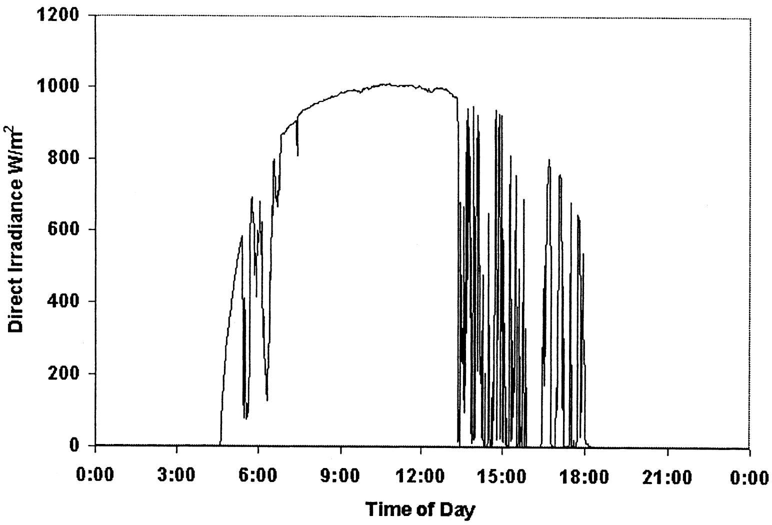

Fig. 9. Measured 1-min direct normal irradiance data for 1 June 1999 for comparison with Figs. 7 and 8.

of their simulations, models and theories rather than on the peculiarities of aggregated data sets. The algorithm produces smooth time series data preserving original data averages. Constraints may be used to bound the interpolated data within user-specified limits. Such smoothly interpolated data may be helpful in evaluating the performance of renewable energy systems at various time series resolutions.

Acknowledgements—This work was performed under DOE contract no. DE-AC-36-98-GO10337 within the Center for Renewable Energy Resources at the National Renewable Energy Laboratory.

REFERENCES

Bouhaddou H., Hassani M. M. and Zeroual A. (1997) Stochastic simulation of weather data using higher order statistics. *Renew. Energy* **12**(1), 21–37.

Diggle P. J. (1990). *Time Series, A Biostatistical Introduction*, Oxford Science, Clarendon Press.

Epstein E. S. (1991) On obtaining daily climatological values from monthly means. *J. Climate* **4**, 365–367.

Gelegenis J. J. (1999) Estimation of hourly temperature data from their monthly average values: case study of Greece. *Renew. Energy* **18**(1), 49–60.

Gueymard C. (1994). *Solar Energy* **53**(1), 57–71.

Hamming R. W. (1986). *Numerical Methods for Scientists and Engineers*, Dover Publications.

Knight K. M., Klein S. A. and Duffie J. A. (1990) A methodology for the synthesis of hourly weather data. In *Proceedings of Solar 90 Annual Conference of the American Solar Energy Society*. *Mar 19–22, Austin, TX*.

Mora-Lopez L. and Sidrach-de-Cardona M. (1998) Multiplicative ARMA models to generate hourly series of global irradiation. *Solar Energy* **63**(5), 283–291.

Noble B. and Daniel J. W. (1988). In 3rd edn, Applied Linear Algebra, pp. 374–376, Prentice-Hall, Englewood Cliffs, NJ.