JACOBI SVD ALGORITHMS FOR TRACKING OF NONSTATIONARY SIGNALS

F. Lorenzelli and K. Yao

Electrical Engineering Department
University of California
Los Angeles, CA 90024-1594
lorenz@ee.ucla.edu, yao@ee.ucla.edu

ABSTRACT

In this paper we consider the algorithm for SVD updating based on Jacobi rotations. In order to overcome the trade-off between accuracy and updating rate intrinsic in the original algorithm, we propose two schemes which improve the overall performance when the rate of change of the data is high. In the "variable rotational rate" scheme, the number of Jacobi rotations per update is dynamically determined. In the "variable forgetting factor" approach, the effective width of the observation adjusts to the data nonstationarity. The former scheme ensures closeness to convergence at all times, while the latter adapts the response to data variation. We consider applications of the SVD updating algorithm to speech processing of segmentation, adaptive parameter estimation, and glottal closure detection.

1. INTRODUCTION

In numerous multi-channel estimation, filtering, and beamformation signal processing applications, the received waveforms are often used in a matrix format. Various linear algebraic and optimization techniques are employed to extract from these matrices useful information for further processing. Basic parameters such as numerical rank and singular values/vectors/subspaces are used for the calculation of filter weights, predictor coefficients, parametrical spectral estimators, etc. In this paper, we investigate the algorithmic and architectural relationships among the input update rate, the rate of convergence of the Jacobi-SVD algorithm, and the quality of the SVD processed outputs. This approach provides new insights on the selection of forgetting factors needed in adaptive signal processing. We obtain a real-time nonstationarity indicator of the observed data in terms of their singular value behavior. This indicator comes without additional computational expense, as a regular part of the Jacobi-SVD algorithm for real-time processing. We demonstrate its usefulness to problems in speech segmentation and linear prediction.

2. THE JACOBI SVD ALGORITHM

If the SVD of the data matrix, $Y_m = U_m \Sigma_m V_m^{\mathrm{H}}$, is known at time m, the updated matrix is given by

Jacobi SVD updating algorithm

Given the computed matrices U_m, Σ_m, V_m

- application of forgetting and vector projection
  $\Sigma'_m \leftarrow \beta \Sigma_m;\ x'_{m+1} \leftarrow x_{m+1} V_m$
  - QR updating
  $$\begin{pmatrix} \Sigma'_m \\ 0 \end{pmatrix} \leftarrow Q_{m+1}^{\mathrm{H}} \begin{pmatrix} \Sigma'_m \\ x'_{m+1} \end{pmatrix}$$
  $$U'_{m+1} \leftarrow \begin{pmatrix} U_m & 0 \\ 0 & 1 \end{pmatrix} Q_{m+1}$$

$Q_{m+1} \in \mathbf{C}^{(n+1)\times(n+1)}$, orthogonal

- Jacobi rotations (rediagonalization)
  $U_{m+1} \leftarrow U'_{m+1}, \quad \Sigma_{m+1} \leftarrow \Sigma'_m, \quad V_{m+1} \leftarrow V_m$
  for $k = 1, \ldots, \ell$; for $i = 1, \ldots, n-1$
  Apply a Jacobi rotation to rows and columns i and $i+1$ of Σ_{m+1}
  Propagate the rotations to U_{m+1} and V_{m+1}
  end

It is known that both the QR and the Jacobi rotation steps can be implemented on a parallel/systolic architecture. [7] proposed a triangular array composed of relatively simple cells to implement these operations characterized by a fixed forgetting rate and a fixed throughput.

2.1. Variable Rotational Rate Scheme

Recently, we have performed extensive study on the tracking behavior for nonstationary data of the Jacobi SVD algorithm. From [6], it is known that in high SNR scenarios, the tracking error within a given time window is only a function of the actual data variation, for any updating rate, given that the rotations are computed at fixed speed. Moreover, it can be proved that for sufficiently slowly changing data, a slowly updating implementation of the Jacobi SVD algorithm produces the same (or better) estimates than a higher throughput implementation, for equal computational rate [3, 4]. When the data variation increases, it would seem that a higher updating rate is a reasonable choice. However, by increasing the input rate without increasing the computation rate, the computed singular matrices are now farther from convergence. The idea we explored is to "decouple" the updating rate from the speed at which rotations are computed ("rotational rate"). The updating rate is kept constant, while one dynamically varies the computational speed, according to the degree of data nonstationarity. This decoupling can be achieved either by having processors of

0-7803-2431-5/95 $4.00 © 1995 IEEE

varying computational power, or by mapping the algorithm onto variable size networks of processors.

Since we are interested in minimizing the number of rotations, we want to know the initial convergence behavior of the Jacobi algorithm. Unfortunately, no general theory is known to explain this phenomenon. What is known is that *eventually* the off-norm will decrease to zero in a quadratic fashion, but how long it takes to reach this stage cannot be exactly predicted. In parallel realizations of the Jacobi algorithm, where the off-norm monitoring is particularly difficult, it is common practice to set the number of sweeps to a predefined value. This approach usually results in a significant number of unnecessary sweeps. Recently, Götze proposed a novel criterion for monitoring the stage of the diagonalization, which is very appealing for parallel implementations [1]. Nevertheless, this criterion is only useful to *detect* convergence, and not to *predict* when it might occur.

It can be shown that the mere presence of noise (and no change in the characteristics of the desired signals) can cause perturbations in the singular subspaces, but this variation is quite limited in high SNR. On the other hand, large changes in the data may require a larger amount of computations for the singular matrices computation, compared to smaller variations [4]. We limit ourselves to the cases in which the amount of perturbation can be considered as a monotonic function of the rate of data change. Consider the QR factorization required by the updating algorithm, where Σ'_m is upper triangular. Since we deal here with continuous variations of the signal subspace, it is likely that the $r \times r$ leading submatrix of Σ'_m will always contain large entries above diagonal (r is the numerical rank of the data matrix). In order to give an estimate to the number of Jacobi rotations needed to diagonalize Σ'_m, what is of interest is the amount of fill-in in the submatrix of Σ'_m composed of its $(n-r)$ rightmost columns. This is in turn determined by the norm of $x_{m+1}V_m^{\mathrm{n}}$. Given the partitioned right singular matrix, $V_m = (V_m^{\mathrm{s}}, V_m^{\mathrm{n}})$, and the incoming vector x_{m+1}, define $\alpha_{m+1} \equiv \|x_{m+1}V_m^{\mathrm{n}}\|/\|x_{m+1}\|$, $m = 0, 1, \ldots$ We choose the quantity α_{m+1} to represent the degree of nonstationarity of the incoming data. This assumption is based on the fact that, in environments characterized by a high signal-to-noise ratio (SNR), and where there is little or no change in the signal parameters, then $\alpha_{m+1} \approx 0$. Remember that even in complete stationarity, the mere presence of noise makes α_{m+1} different from zero. The indicator α_{m+1} is easily computed in the Jacobi algorithm. It can be shown that this indicator is, at least for not too large variations, a monotonic function of the data change.

We have subsequently analyzed the initial convergence behavior of the Jacobi SVD algorithm, trying to determine the relationship between the number of Jacobi rotations required for a satisfactory level of convergence (parameter ℓ of the Jacobi algorithm above) and the degree of data variation (indicator α_{m+1}). For this reason, we considered the behavior of the off-norm of the matrix Σ'_m as function of ℓ. We concluded the following [4]: 1) When α_{m+1} is close to zero, the fill-in concentrates in the r rows of Σ'_m corresponding to the r larger singular values. Therefore, a substantial decrease of the off-norm is achieved only for $\ell \geq r$. 2) If α_{m+1} is substantially different from zero, but $\alpha \ll 1$, then the fill-in spreads throughout the matrix Σ'_m, and the off-norm decreases more slowly as ℓ increases from 1 to n. 3) When $\alpha \approx 1$, then the number of Jacobi rotations required for diagonality can become considerably smaller than for smaller α_{m+1}.

We subsequently considered the behavior of the off-norm of Σ'_m in time, in the context of periodic updating. If $\ell = 1$, fixed, then the off-norm stops decreasing after a given value, and assumes a pseudo-periodic pattern of period n. This phenomenon is due to the use of *permuted* Jacobi rotations, and the variable balance between the off-norm increase caused by the QR updating step on the one hand, and the off-norm decrease achieved by the Jacobi rotations on the other.

For all the previous considerations, we propose the following variable rotational rate scheme, for medium to high SNR, noise power σ_{N}^2, and numerical rank $= r$:

- Compare the nonstationarity indicator to a threshold μ. Since in a totally stationary environment the value of $\|x_{m+1}V_m^{\mathrm{n}}\|$ approximates, on average, the value $\mu_0 \equiv \sigma_{\mathrm{N}}\sqrt{n-r}$, we suggest that the threshold μ be a multiple of μ_0.
- If $\alpha_{m+1} \leq \mu$, then choose a value for ℓ not smaller than r. Otherwise choose $\ell \geq n$.
- Choose a high enough forgetting factor, which guarantees that the diagonal elements of Σ_m are sufficiently large (cf. next section).

2.2. Variable Forgetting Scheme

We have also studied the relationship between the data variation an the appropriate choice for the forgetting factor. It can be shown analytically, that SVD tracking requires narrower observation windows, as the rate of data variation increases. In particular, it can be shown that a rate of change q times faster in the signal parameters calls for an observation window q times shorter, which is achieved by reducing the forgetting factor, β, to β^q [4]. The aperture of the effective exponential window is controlled by the forgetting factor in a very simple fashion. No special hardware is required, other than a device which dynamically updates the value of the forgetting factor, such that within the associated observation window the data are approximately stationary. During processing, the forgetting rate, β, can be determined by equating the window width (function of β) with the observed stationarity interval, N_{w}. According to our analysis, we choose the parameter β so that, *within the observation window*, the change in the data is smaller than a given threshold. Given a value for the forgetting factor, β, and a threshold, b, the length of the observation window, N_{w}, is defined so that $\beta^{N_{\mathrm{w}}} = b$, i.e., $N_{\mathrm{w}} = \log b/\log\beta$. Consistently with our analysis, the data change from sample m to sample $(m+1)$ is measured by the quantity $\|x_{m+1} - x_m\|^2 = \|(x_{m+1} - x_m)V_m^{\mathrm{s}}\|^2 + \alpha_{m+1}^2$. When the data are stationary, $\alpha_{m+1} \approx \sigma_{\mathrm{N}}$. If the variations are assumed identical for N_{w} consecutive samples, then, given a bound, Ξ, on the allowable data change, one can compute N_{w} as $N_{\mathrm{w}} = \Xi/\|x_{m+1} - x_m\|^2$. If one specifies the data generation model, then all the terms involved can be given as function of the signal and noise parameters. The adaptation of the forgetting factor is possible only within given bounds. If it is required that the number of Jacobi rotations which rediagonalize Σ' be kept low, then the amount of fill-in produced by the QRD step has to be limited. The fill-in is dependent on the relative magnitude of the elements of x'_{m+1} and the diagonal entries of Σ'. The size of the diagonal elements of Σ is in turn related to β according to $1/\sqrt{1-\beta^2}$. By assuming that the incoming data vectors have constant norm (on average, at least locally), then a rough estimate of $\beta_{\min}$ is given by satisfying $\beta/\sqrt{1-\beta^2} \gg 1$, i.e., $\beta \gg \beta_{\min} = 0.72$. In conclusion, the forgetting factor can be decreased but not beyond a given value. The proposed variable forgetting scheme can be summarized as follows:

- Determine at every time instant the duration of the stationarity window, N_w;
- Given a threshold b, compute β so that $\beta^{N_w} \leq b$;
- Make sure that β is not too small, $\beta \geq \beta_{\min}$.
- compute β as $\beta = \max\left\{b^{1/N_w}, \beta_{\min}\right\}$.

3. NUMERICAL EXAMPLES

The proposed SVD updating algorithm can find application in many situations, such as beamforming, adaptive filtering, DOA tracking. In this section, we wish to show how the same algorithm can be applied to the speech processing scenario.

In the first example, we demonstrate the real-time tracking of the SVD of the data matrix of speech signals. Speech utterances can be segmented into *voiced* and *unvoiced* sounds. A widely used speech production model is composed of an excitation, of periodic pulses for voiced speech or white noise for unvoiced speech, onto a vocal tract filter modelled by a p-th order all-pole time-varying linear filter. Construct a $L \times J$ Toeplitz data matrix Y_m, with the first row given by $\{y(m-L+1), y(m-L), \cdots, y(m-J-L+2)\}$, constructed from an observed speech waveform $y(m)$ corresponding to the utterance "test." The SVD of Y_m is then dependent on both the coefficients of the all-pole filter, and on the correlation properties of the excitation. [2] showed that the rank of this matrix is related to the model order p. When the excitation is periodic, then the singular value distribution of Y_m tends to display a low-rank structure, and the smallest singular values are close to zero; when the excited is by random white noise, then the singular value distribution spreads out. Often the voiced sounds are associated with higher energy than unvoiced ones. Consider the singular values corresponding to a 20×20 data matrices associated with the utterance of the English word "test" by a male speaker. In Fig. 1 the distributions are normalized to the first singular value, and we observed that the singular value distribution for the sound /e/ falls off more rapidly than that of the sound /s/. The previous observations motivate the use of the proposed scheme, which tracks the behavior of the data matrix's singular values and right singular vectors. The nonstationary indicator α_{m+1} can be used to detect the boundary between voiced and unvoiced sounds, task which is known to be difficult in practice. If the sound is voiced, then α_{m+1} falls below a threshold value, whereas it remains above it when the sound is unvoiced. These results are shown in Fig. 3, where the waveform associated to the utterance "test" and the indicator α_{m+1} are plotted against time. When the voiced sound /e/ exists over time unit of about 1,500 to over 3,000, there is a corresponding dip in the indicator over this interval.

In the second example, we show how the computed SVD can be used for adaptive parameter estimation. Consider the running SVD of the data matrix, and compute an estimate of the all-pole filter coefficients. For an AR filter model, $y(m) = Gu(m) + a_1 y(m-1) + \cdots + a_p y(m-p)$, for all time instants m, where $u(m)$ are unit variance, zero mean excitation samples. Given the data matrix, Y_m, defined earlier, when $J > p$, and the $J \times 1$ vector $a \equiv (1, a_1, \ldots, a_p, 0, \ldots, 0)^{\mathrm{T}}$, we have that $Y_m a = G u_m$, where $u_m \equiv (u(m-L+1), u(m-L+2), \ldots, u(m))$. In order to estimate the AR coefficients, a possible strategy is to find the vector $\hat{a}$ that minimizes the norm of $Y_m a$. If the SVD of Y_m is given as $Y_m = U_m \Sigma_m V_m^{\mathrm{H}}$, then we have $\hat{a} = \arg\min_{x, x(1)=1} \|Y_m x\| = \arg\min_{x,x(1)=1} \|\Sigma_m V_m^{\mathrm{H}} x\|$. The vector $\hat{z}$ which minimizes the norm of $\Sigma_m z$ is given by $\hat{z} = e_n^{\mathrm{T}} \equiv (0, 0 \ldots, 0, 1)^{\mathrm{T}}$, and the norm is equal to the smallest singular value. It follows that $\hat{a} = v_{\min}/v_{\min}(1)$, where $v_{\min}$ is the right singular vector corresponding to the smallest singular value. In conclusion, the desired estimated AR coefficients are given by the singular vector corresponding to the smallest singular value. In Fig. 2 we show the coefficient values of a order-9 AR filter, estimated adaptively by use of the proposed SVD updating algorithm. The coefficient shown were computed between times 2160 and 2165 (corresponding to sound /e/, between glottal closures), and between times 4360 and 4365 (corresponding to sound /s/)

It is well-known that linear prediction models fit best during speech segments of less than one pitch period, between instants of glottal closure, or "epochs." At glottal closure, the excitation is present in the data, with the consequence that the linear prediction model does not fit the data properly and the prediction error is large. The large deviation between actual and predicted data around a glottal closure instant, due to the abrupt change in glottal flow, has been used for epoch estimation by many authors. Recently, a new method for glottal closure detection has been proposed, based on the recursive computation of the Frobenius norm on a sliding window [5]. The algorithm for SVD updating proposed here can be successfully employed for glottal closure detection, by keeping track of the behavior of the individual singular values. In Fig. 4 we show the behavior of three computed singular values, namely $\sigma_{\max}$, $\sigma_{\min}$, and σ_5, corresponding to the sounds /e/ and /s/ of the utterance "test." From this figure we note the clear periodic pattern of the singular values during a voiced segment, with pulses at the glottal closure instants. No particular pattern is discernible for the unvoiced sound. For comparison, we also show the running Frobenius norm for the two cases. The computational complexity required here is much higher than for the method of [5], but the information extracted is also larger, and can be used for other purposes (detection of voiced-unvoiced segments, adaptive parameter estimation, etc.). Note also that the smaller singular values have sharper peaks around the epochs than both the larger singular values and the Frobenius norm.

4. REFERENCES

[1] J. Götze. "Monitoring the Stage of Diagonalization in Jacobi-Type Methods". *Proc. of ICASSP*, 1994.

[2] K. Konstantinides and K. Yao. "Stat. Analy. of Eff. SV in Matrix Rank Determination". *IEEE Trans. on Acous.,Speech, and Signal Proc.*, May 1988.

[3] F. Lorenzelli and K. Yao. "SVD Updating for Nonstationary Data". *IEEE VLSI Signal Proc.*, 1994.

[4] F. Lorenzelli and K. Yao. "Updating the Jacobi SVD for Nonstationary Data". Submitted for publication

[5] C. Ma, Y. Kamp, and L.F. Willems. "A Frob. Norm App. to Glottal Closure Detection from the Speech Signal". *IEEE Trans. Speech and Audio Proc.*, April 1994.

[6] M. Moonen, P. Van Dooren, and J. Vandewalle. "A SVD Updating Algorithm for Subspace Tracking". *SIAM J. Matr. Anal. and Appl.*, October 1992.

[7] M. Moonen, P. Van Dooren, and J. Vandewalle. "A Systolic Array for SVD Updating". *SIAM J. Matr. Anal. and Appl.*, April 1993.

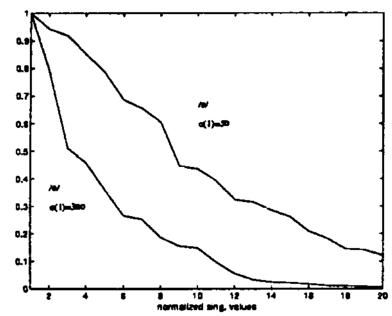

Figure 1: Normalized singular value distributions corresponding to sounds /e/ and /s/

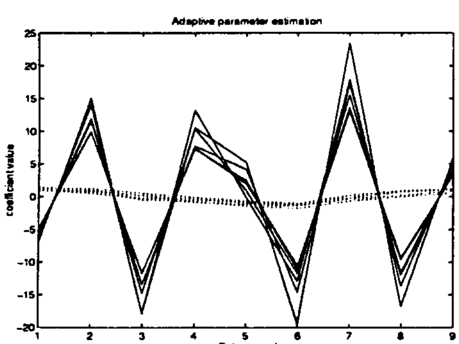

Figure 2: AR coefficient of order-9 filter computed using the SVD updating algorithm. Solid lines: times 2160 to 2165, sound /e/ between glottal closures. Dotted lines: times 4360 to 4365, sound /s/

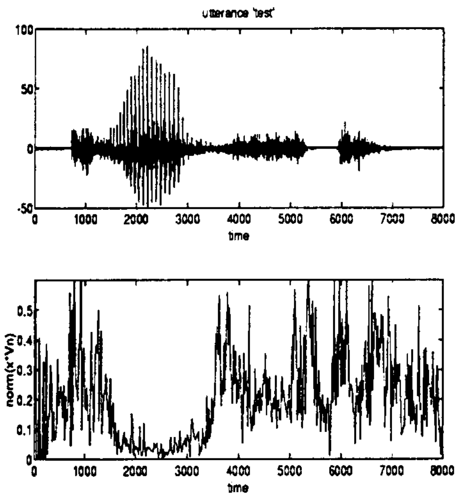

Figure 3: Time waveform for the word "test" and norm of $x_{m+1}V_m^n$ as function of time

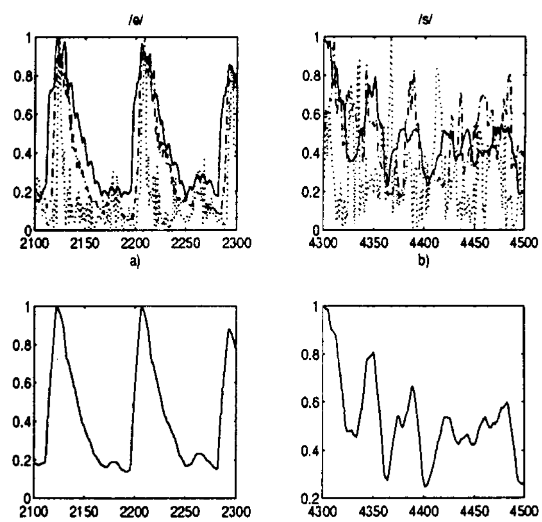

Figure 4: Normalized singular values ($\sigma_{\max}$ (solid), σ_5 (dash-dot), and $\sigma_{\min}$ (dotted)) and Frobenius norm of windowed data matrices. a) and c) normal. sing. values for segment corresponding to /e/ and /s/; b) and d) normal. Frobenius norm for segment corresponding to /e/ and /s/