A NEW TECHNIQUE TO FILTER REDUCTION FOR SPEECH SIGNAL PROCESSING SYSTEMS

Luowen Li , Lihua Xie , Gang Li , Yeng Chai Soh

School of Electrical and Electronic Engineering
Nanyang Technological University
Nanyang Avenue, Singapore 639798

ABSTRACT

In many applications, one needs to approximate a filter of very high order with that of lower order. To reduce the order of the filter, some techniques such as balanced model reduction approach are often applied. In this paper, we will introduce a new technique which is based on minimizing the H_2-norm between the filter of very high order and the reduced one. This technique shows much better performance than other existing model reduction methods and is applied to estimating the vocal tract filter for speech processing systems. A speech processing example is presented to demonstrate the design procedure and the performance of the proposed algorithm.

1. INTRODUCTION

In parametric approach to speech analysis/synthesis, speech signals are usually modeled as the output of a filter, called vocal tract filter, excited by a certain input. This filter is traditionally given by a transfer function of all-pole [1]. The main advantage of using all-pole filter is that the corresponding filter parameters can be obtained easily. To have a very accurate modeling of the vocal tract, a more general pole-zero filter has to be used. In fact, the mechanism of speech production suggests that the vocal tract be modeled with a pole-zero transfer function (see, e.g., [2]). The main reason why the general pole-zero filter has not been popularly used in vocal tract modeling is due to the difficulty in estimating its parameters with the only available original speech signal.

In this paper, we present a new method for estimation the parameters of the pole-zero vocal tract filter. The basic idea is first to model the vocal tract with an all-pole filter of very high order, then to convert it into a pole-zero filter of very low order using one of the model reduction techniques. By doing so, high quality synthetic speech can be achieved with very low bit rate coding. The main objective in this paper is to develop a new efficient model reduction technique for estimating the vocal tract filter of pole-zero type.

Model/filter reduction has been an important research topic during the past few decades and there has been a well-developed theoretical foundation for this problem under a variety of approximation criteria, for example, the Balanced truncation technique(BT) [3] [4], Hankel-norm approximation approach [5], Impulse response gramin(IRG) [6] and Covariance equivalent reduction methods [8] [9]. The well known effective balanced truncation technique is widely used both in model reduction and filter reduction [10] [11] [3]. The basic idea of this technique is to find a similarity transformation and identify the weak modes of the system which have relatively little affect on the system and then truncate them. Recently, this technique has been used in approximating an linear phase FIR filter by an IIR one [6] [7]. In this paper, we will introduce the gradient flow(GF) approach into the filter reduction for the speech signal production system. This is an optimality-based approach and the primary principle is to minimize H_2-norm of the discrepancy between the original filter and the reduced one [12]. We will develop both continuous and iterative algorithms and prove the convergence of them. Here, we represent the filters in state-space forms and treat the minimization problem over a subclass of stable reduced-order filters parameterized by a projection matrix instead of the whole class of all the reduced-order filters. The restriction to this subclass enables us to avoid the stability constraint entirely and leads to a tractable minimization problem over the compact Stiefel manifold. In addition, the local minimum is guaranteed to exist over the subclass. A practical example is implemented to demonstrate the effectiveness of the proposed technique.

2. PROBLEM FORMULATION

A time-invariable linear digital filter, say the all-pole vocal tract filter of high order obtained with a lattice estimator, can be represented by a state-space realization:

$$x(k+1) = Ax(k) + Bu(k) \tag{2.1}$$

$$y(k) = Cx(k) \tag{2.2}$$

where $x(k) \in R^n$ is the state, $y(k) \in R^q$ is the output, $u(k) \in R^p$ is the input. A, B, C are of suitable dimensions.

A reduced order filter is given (say m-th order with $m < n$):

$$x_m(k+1) = A_m x_m(k) + B_m y(k) \tag{2.3}$$

$$y_m(k) = C_m x_m(k) \tag{2.4}$$

where $A_m \in R^{m\times m}$, $B_m \in R^{m\times q}$, $C_m \in R^{p\times m}$. The mismatch between the full-order $G(z)$ and the reduced-order $G_m(z)$ will be measured by the square of H_2 norm of their difference $G_e(z)$, i.e.,

$$\|G_e(z)\|_2^2 = \|G(z) - G_m(z)\|_2^2 \tag{2.5}$$

which is often termed as the quadratic model-reduction cost.

Note that one state-space realization $(A_e\ , B_e\ , C_e)$ of the error model $G_e(z)$ is given by

$$(A_e, B_e, C_e) = \left(\begin{bmatrix} A & 0 \\ 0 & A_m \end{bmatrix}, \begin{bmatrix} B \\ B_m \end{bmatrix}, \begin{bmatrix} C & -C_m \end{bmatrix} \right) \tag{2.6}$$

Then it is a standard fact that the cost can be conveniently expressed in terms of the controllability Gramian P and observability Gramian Q of this realization. Namely, there holds

$$\|G_e(z)\|_2^2 = J(A_m, B_m, C_m) = tr(C_e P C_e^T) = tr(B_e^T Q B_e) \tag{2.7}$$

where P and Q satisfy

$$A_e P A_e^T - P + B_e B_e^T = 0 \tag{2.8}$$

$$A_e^T Q A_e - Q + C_e^T C_e = 0 \tag{2.9}$$

It has been proven that any minimizing solution (A_m, B_m, C_m) must be of the form [13]

$$(A_m, B_m, C_m) = (TAV, TB, CV) \tag{2.10}$$

where $V \in R^{n\times m}$ and $T \in R^{m\times n}$ satisfy $TV = I$.

Hence, the original filter reduction problem amounts to minimizing $J(TAV, TB, CV)$ with respect to $(T, V) \in R^{m\times n} \times R^{n\times m}$ subject to the two constraints

$$(i) TV = I \qquad (ii) TAV \; is \; stable \tag{2.11}$$

To make this reduction problem more tractable, it has been proven in [14] that the above problem can be modified as the following minimization problem over a much smaller set:

$$(A_m, B_m, C_m) = \{(U^T AU, U^T B, CU) | U \in St(m, n)\} \tag{2.12}$$

where $U^T AU$ is stable and $St(m, n)$ is the so-called Stiefel manifold defined by

$$St(m, n) = \{U \in R^{n\times m} | U^T U = I\} \tag{2.13}$$

Hence, we have:

$$J(A_m, B_m, C_m) = J(U) = J(U^T AU, U^T B, CU) \tag{2.14}$$

The minimization over the latter smaller model set will lead to a local minimum more quickly and the associated computation will be less expensive.

3. GRADIENT FLOW TECHNIQUE

It is clearly known from Section 2 that our filter reduction objective is formulated to find a transformation matrix U such that the cost $J(U)$ is minimized. To make a more explicit formula for $J(U)$, we partition the solutions P and Q of the Lyapunov equation (2.8)-(2.9) as

$$P = \begin{bmatrix} P_{11} & P_{12} \\ P_{12}^T & P_{22} \end{bmatrix} \;\; and \;\; Q = \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{12}^T & Q_{22} \end{bmatrix} \tag{3.1}$$

As a result, the lyapunov equations (2.8)-(2.9) become equivalent to

$$AP_{11}A^T - P_{11} + BB^T = 0 \tag{3.2}$$

$$AP_{12}U^T A^T U - P_{12} + BB^T U = 0 \tag{3.3}$$

$$U^T AUP_{22}U^T A^T U - P_{22} + U^T BB^T U = 0 \tag{3.4}$$

$$A^T Q_{11} A - Q_{11} + C^T C = 0 \tag{3.5}$$

$$A^T Q_{12} U^T AU - Q_{12} - C^T CU = 0 \tag{3.6}$$

$$U^T A^T UQ_{22}U^T AU - Q_{22} + U^T C^T CU = 0 \tag{3.7}$$

and the cost $J(U)$ can be rewritten as

$$\begin{aligned} J(U) &= tr[C^T C(P_{11} + UP_{22}U^T - 2P_{12}U^T)] \\ &= tr[BB^T(Q_{11} + UQ_{22}U^T + 2Q_{12}U^T)] \end{aligned} \tag{3.8}$$

Lemma 3.1. *If P and Q satisfy*

$$EPF^T - P + X = 0 \;\; and \;\; E^T QF - Q + Y = 0 \tag{3.9}$$

then there holds

$$tr(Y^T P) = tr(X^T Q) \tag{3.10}$$

Consider the symmetric characteristic of $R^T(U)U$, we can form the gradient of the cost $J(U)$ as following theorem.

Theorem 3.1. *: For any $U \in St(m, n)$, the gradient of $J(U)$ on $St(m, n)$ is given by*

$$\bigtriangledown J(U) = (I - UU^T)R^T(U)$$

where

$$\begin{aligned} R^T(U) = \; & P_{12}^T(-C^T C + A^T Q_{12}U^T A) + \\ & P_{22}(U^T C^T C + U^T A^T UQ_{22}U^T A) + \\ & Q_{12}^T(BB^T + AP_{12}U^T A) + \\ & Q_{22}(U^T BB^T + U^T AU^T P_{22}U^T A^T) \end{aligned} \tag{3.11}$$

From Theorem 3.1, it is clear that the necessary conditions for the minimality of $J(U)$ in $St(m, n)$ are

$$(I - UU^T)R^T(U) = 0 \; and \; U^T U = I \tag{3.12}$$

Solving the above equation is a difficult problem because there may exist multiple solutions and the derived solution may be a maximum one. Alternatively, we form a gradient flow as follows:

$$\dot{U} = - \bigtriangledown J(U) = (UU^T - I)R^T(U) \tag{3.13}$$

for solving the optimization problem. To be feasible, the optimization must ensure that the solution U of (3.13) exists and moreover remains in the $St(m, n)$ manifold. Indeed, the following theorem provides a positive answer. The proof can be derived similarly as in [12]

Theorem 3.2. *: Let the initial value of the ODE(3.13) be given by $U(0) = U_0 \in St(m, n)$. Then, we have the following results*
(1) The ODE has a unique solution $U(t)$ defined for all $t \geq 0$;
(2) The solution $U(t)$ stays in $St(m, n)$ for all $t \geq 0$;
(3) The cost $J(U)$ is non-increasing along $U(t)$ with $J(U(t_2)) - J(U(t_1)) = -2\int_{t_1}^{t_2} \|(I - UU^T)R^T\|_F^2 dt$ where $\forall t_2 \geq t_1 \geq 0$ and $\|\cdot\|_F$ denotes the Frobenius norm;
(4) There holds

$$\lim_{t\to\infty} \dot{U}(t) = \lim_{t\to\infty} (UU^T R^T - R^T) = 0$$

(5) The solution $U(t)$ converges to component of the set of critical points of $J(U)$;
(6) There exists a time sequence t_k with $t_k \geq 0$ and $\lim_{t\to\infty} t_k = \infty$ such that the corresponding sequence $U(t_k)$ converges to a critical point of $J(U)$.

Although the solution of the ODE (3.13) can lead to a minimum point of the cost, a recursive algorithm is usually preferred in a digital environment. Therefore, in the remainder of this section, we shall see how a special form of the gradient flow can be exploited to yield a recursive algorithm which automatically iterates over the Stiefel manifold and should be much easier to implement with a digital computer.

Since $R^T(U)U$ is symmetric, (3.13) can be rewritten as

$$\dot{U} = UR^TU - RU^TU = (UR^T - RU^T)U \qquad (3.14)$$

The matrix exponential $e^{(UR^T - RU^T)t}$ is orthogonal for any real scalar t as $UR^T - RU^T$ is skew-symmetric. Hence, the following iterative form is suggested:

$$U_{k+1} = e^{t_k(U_k R_k^T - R_k U_k^T)} U_k \qquad (3.15)$$

where R_k is as defined in (3.11) with U replaced by U_k and t_k is the k-step size.

Obviously, all the matrices generated by this algorithm from any starting $U_0 \in St(m,n)$ will remain in $St(m,n)$ no matter how the step t_k is chosen. However, it is important to ensure that the cost $J(U)$ is decreasing by a proper choice of t_k. This is specified in the following theorem.

Theorem 3.3. *[12] There exists a constant c such that with the step-size chosen as $0 < t_k \le c$ and $U_0 \in St(m,n)$,*

$$J(U_{k+1}) \le J(U_k), \quad \forall k = 0, 1, 2, \cdots$$

where the equality holds if and only if a critical point of $J(U)$ is reached.

To sum up, the above two algorithms(Continuous and Iterative) are implemented following the steps below:

1. Obtain the balanced realization (A, B, C) of the higher order signal model;
2. Set $U_0 = U(0) = [I_{r\times r} \ \ 0_{r\times(n-r)}]^T$ and solve the ODE (3.13) or update the iterates (3.15) with an appropriate step size;
3. Form the reduced order filter with the obtained final U from Step 2 as: $(A_m, B_m, C_m) = (U^TAU, U^TB, CU)$.

4. EXPERIMENT RESULT

Now, we present some experimental results. The data file, called 'clean', is standard speech signal obtained from the database of Sheffield University. The signal presents the utterance 'Fred can go, Susan can't go, and Linda is uncertain' spoken by a female with its waveform shown in Figure 1. The sampling frequency is $f_s = 20kHz$. The duration is $3.574sec$, that is 71480 samples.

Figure 2 shows the original speech signal of 400 samples from 8101 to 8500. This is a typical voiced frame. The all-pole vocal tract filter of order 30, denoted by $G(z)$, is obtained using a lattice estimator. This filter is approximated with a pole-zero filter of order 6 using three different algorithms: GF, BT and IRG.

To compare the reduction results, we check the following two error indices:

$$\delta_1 = \|G(z) - G_m(z)\|_2, \ \ \delta_2 = \frac{\|h - h_r\|_\infty}{\|h\|_\infty}$$

where h and h_r are the impulse responses of the high order filter and the reduced-order ones, respectively. The comparisons are shown in Table 1:

	GF	BT	IRG
δ_1	5.0407	5.0444	16.1653
δ_2	3.9104e-3	3.9230e-3	2.7913e-3

Table 1: Error comparison of existed techniques

Figure 3 and 4 depict the synthetic speech with $G(z)$ and $G_{GF}(z)$ obtained with GF algorithm, both excited with the same prediction error signal of the lattice estimator.

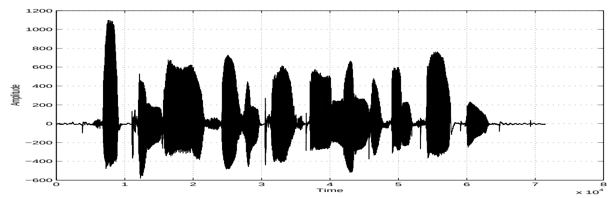

Figure 1: The waveform of the sample speech signal.

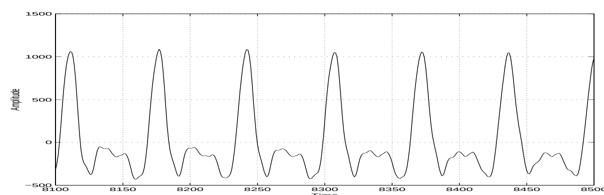

Figure 2: One typical voiced frame of the speech signal.

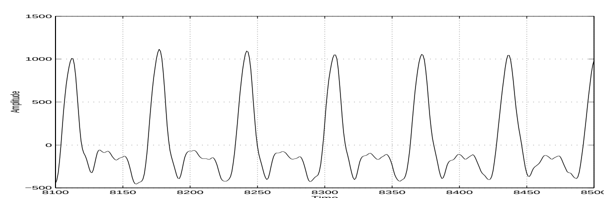

Figure 3: The synthetic signal with $G(z)$.

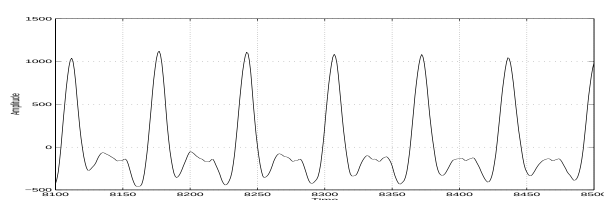

Figure 4: The synthetic signal with $G_{GF}(z)$.

Clearly, $G_{GF}(z)$ yields almost the same waveform with only 13 parameters while $G(z)$ requires 30 parameters for implementation.

Furthermore, we consider another typical unvoiced frame with 350 samples from 151 to 500. It is shown in Figure 5.

Once again, the three algorithms are appliced for model reduction. The corresponding error comparison is shown in Table 2:

	GF	BT	IRG
δ_1	0.1554	0.1606	0.2195
δ_2	9.5603e-3	9.6034e-3	9.7346e-3

Table 2: Error comparison of existed techniques

Figure 6 and Figure 7 are the synthetic frame obtained by a 20th order all-pole filter and a 5th-order ROF via GF approach.

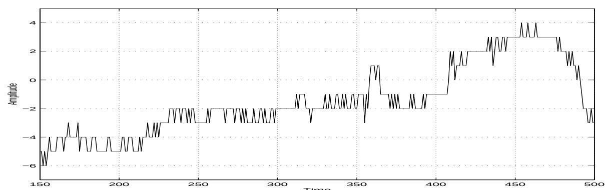

Figure 5: One typical unvoiced frame of the speech signal.

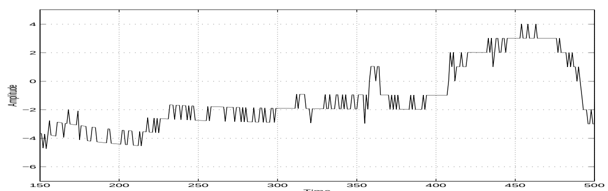

Figure 6: The synthetic signal by the high order filter.

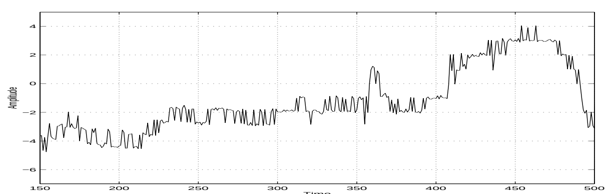

Figure 7: The synthetic signal by the ROF.

5. CONCLUSION

In this paper, we have developed the so-called gradient flow approach and applied it to vocal tract filter estimation for speech signal processing systems. Through a practical speech signal experiment, we found that the reduction result obtained via the gradient flow approach has better performance comparison with those obtained by other techniques such as BT and IRG.

6. REFERENCES

[1] C. Wheddon and R. Linggard, "Speech and language processing", *Chapman and Hall*, 1990.

[2] S. Yim, D. Sen, and W.H. Holmes, "Comparison of ARMA modeling methods for low bit rate speech coding", *Proc. Int. Conf. on Acoustic., Speech and Signal Processing*, vol. I , pp. 273–276, 1994.

[3] B.C. Moore, "Principal component analysis in linear systems:controllability, observability and model reduction", *IEEE Transactions on Automatic Control*, vol. 26, pp. 17–32, 1981.

[4] L. Pernebo and L.M. Silverman, "Model reduction via balanced state space representation", *IEEE Transactions on Automatic Control*, vol. 27, pp. 382–387, 1982.

[5] S.C. Peng, B.S. Chen, and B.W. Chiou, "IIR filter design via optimal Hankel-norm approximation", *Proc.Inst.Elec.Eng.*, vol. 139, pp. 586–590, 1992.

[6] S. Holford and P. Agathoklis, "The use of model reduction techniques for designing IIR filters with linear phase in the passband", *IEEE Transactions on Singal Processing*, vol. 44, pp. 2396–2404, 1996.

[7] M. Rudko, "a note on th 'approximation of FIR by IIR digital filters: an algorithm based on balanced model reduction'", *IEEE Transactions on Automatic Control*, vol. 33, pp. 687–691, 1988.

[8] R.E. Skelton and B. D. O. Anderson, "Weighted q-markov covariance equivalent realizations", *International Journal of Control*, vol. 49, pp. 1755–1771, 1989.

[9] Z. Wang, L. Li, and H. Unbehauen, "A new approach to model reduction by equivalent realization of second-order information", *Advances in Systems Science and Applications*, pp. 87–92, 1996.

[10] B. Beliczynski, I. Kale, and G.D. Cain, "Approximation of FIR and IIR digital filters:an algorithm based on balanced model reduction", *IEEE Transactions on Singal Processing*, vol. 40, pp. 532–542, 1992.

[11] Ubaid M. Al-saggaf and Gene F. Franklin, "Model reduction via balabced realizations:an extension and frequency weighting techniques", *IEEE Transactions on Singal Processing*, vol. 43, pp. 314–317, 1995.

[12] L. Xie, W. Yan, and Y.C. Soh, "L_2 optimal filter reduction:a close-loop approach", *IEEE Transactions on Singal Processing*, vol. 46, pp. 11–20, 1998.

[13] D.C. Hyland and D.S. Bernstein, "The optimal projection equation and the relationships among the methods of wilson,skelton and moore", *IEEE Transactions on Automatic Control*, vol. 30, pp. 1201–1211, 1985.

[14] W. Yan and J. Lam, "L_2 optimal model reduction", *Proc. the 35th IEEE Conf. Decsion and Control*, pp. 2008–2013, 1996.