

UNDERSTANDING THE EXPLOSIVE DIVERGENCE OF THE FTF ALGORITHM

J. R. Bunch

University of California, San Diego
Department of Mathematics
La Jolla, CA 92093-0112, U.S.A.
jbunch@ucsd.edu

R. C. Le Borne

Tennessee Technological University
Department of Mathematics, Box 5054
Cookeville, TN 38505, U.S.A.
rleborne@tntech.edu

I. K. Proudler

QinetiQ, St. Andrews Road
Malvern Worcs. WR14 3PS, U.K.
i.proudler@signal.qinetiq.com

ABSTRACT

Along with its many desirable properties the Fast Transversal Filter (FTF) algorithm suffers from explosive divergence. This type of divergence occurs when the algorithm is seemingly performing its operations normally, producing usable solutions, when the algorithm appears to suddenly produce extremely large errors and an obviously useless solution. Although it is known that a loss of backward consistency is the cause for the resultant perturbations, i.e., a violation to interrelationships between update parameters are not explicitly enforced by the update equations, it is not known why the algorithm suffers explosive divergence rather than a divergence that grows as a continuous function over time. Algorithms have been proposed to circumvent this problem but it remains to be shown through theoretical justification whether these algorithms have remedied the problem or only put it off to some later iteration. Here, we provide a rationale to explain the explosive character of divergence that is inherent to the manner in which the FTF algorithm is derived.

1. INTRODUCTION

The Fast Transversal Filter (FTF) algorithm is well-known to suffer from a kind of computational divergence that can occur over a very small number of update iterations. We term this divergence *explosive divergence*. Lin [1] as well as Cioffi and Kailath [2], determined that shortly before such divergence one of the update parameters, the so-called conversion factor, attains a computed value that violates a theoretical bound. Independently, Slock and Regalia in [3] and [4], respectively, discussed this phenomenon in terms of a stability domain (or, equivalently, a stability manifold) and attributed it to the loss of backward consistency of the filter's update parameters. Essentially, the FTF algorithm loses its least squares character when constraints that theoretically link the algorithm's parameters, and which are not monitored during the execution of the algorithm, become violated. This motivated what now is termed the *Stabilized* FTF algorithm given by Slock and Kailath in [5]. However, the term *stable* refers to experiments in which the modified algorithm continued for many iterations after the FTF algo-

rithm suffered divergence. It remains to be shown analytically whether this observance is indeed general or if it only holds true for some class of problems. Recently, Bunch, et. al. [6] used the concept of consistency to better predict forthcoming divergence before the FTF algorithm loses its least squares character. However, it remains an open question as to why the divergence of the FTF algorithm is explosive in nature and not a type of divergence that can be modelled by some continuous, increasing function. If it can be shown that the explosive divergence is inherent to the manner in which the FTF was derived, then one must be very careful with using any algorithm that is the result from modifications to the FTF algorithm. We demonstrate that conditions exist in which a subproblem used in the derivation of the FTF algorithm can become ill-posed. We present conditions in which this occurs and mention how these conditions can easily be satisfied with the FTF algorithm. Additionally, this is of interest since it involves a condition that may be inherent to other algorithms.

2. BACKGROUND

The FTF algorithm can be derived by considering four transversal filters responding to a common input [7, ch. 16]. Equivalently, these four filters can be written as least squares problems related in that only the right hand side vector changes. The FTF filter algorithm exploits this relationship through its update equations to yield values that in exact arithmetic are equivalent to the values given from the four transversal filters (or least squares problems). As discussed in Bunch, et., al. [8], the sensitivity to perturbations in these parameters are defined in part by the underlying problem: In this case, the problem would be one or more of the four transversal filters (least squares problems) which serve as a first step in the derivation of the FTF algorithm. In particular, one of these filters (choice of right hand side vector if interpreting this as a least squares problem) yields as its output parameter a column of the pseudo-inverse of its data matrix. Before detailing this, we introduce the notation that is to be followed.

To denote vectors, matrices and real scalars we will

use characters in boldface, capital and lower case, respectively. We define $\mathbf{u}_M(n)$ to represent the $M \times 1$ vector of input measurements $u(i)$, $i = (n - M + 1), (n - M + 2), \dots, n$ that are linearly combined to approximate some response $d(n)$ at time index n . In general,

$$\mathbf{u}_M(i) = [u(i) \ u(i-1) \ \dots \ u(i-M+1)]^t, \quad i = 1, \dots, n.$$

Then the residual in estimating $d(n)$ using $\mathbf{u}_M(\mathbf{n})$ is given by $e(n)$,

$$e(n) = d(n) - \sum_{i=1}^M k_i u(n - i + 1).$$

Here the coefficient vector $\mathbf{k}_M(n) = [k_1 \ k_2 \ \dots \ k_M]^t$ is found by minimizing the cost function

$$\sum_{i=0}^n |e(i)|^2.$$

To fully describe the particular transversal filter of interest, we assemble the vectors $\mathbf{u}_M(i)$ into a matrix A ,

$$A = \begin{pmatrix} \mathbf{u}_M(\mathbf{n})^t \\ \mathbf{u}_M(\mathbf{n}-1)^t \\ \vdots \\ \mathbf{u}_M(\mathbf{M})^t \end{pmatrix} \in \mathbb{R}^{(n-M+1) \times M}.$$

The vector of desired responses

$$\mathbf{d}(n) = [d(n) \ d(n-1) \ \dots \ d(n-M)]^t$$

will be defined through the special choice

$$\mathbf{d}(n) = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} =: \mathbf{1}_{((n-M+1),1)}(n). \quad (1)$$

With (1), the problem for finding the minimizing $\mathbf{k}_M(n)$ is restated now as

$$\min_{\mathbf{k}} \|\mathbf{d}(n) - A\mathbf{k}\|_2 = \|\mathbf{d}(n) - A\mathbf{k}_M(n)\|_2. \quad (2)$$

For A having full rank and letting $\Phi_M(n) = A^t A$, we have

$$\mathbf{k}_M(n) = \Phi_M(n)^{-1} \mathbf{u}_M(n). \quad (3)$$

For least squares applications in signal processing, $\mathbf{k}_M(n)$ is the extended gain vector [7, pp. 578-579]. However, given the special choice (1) for $\mathbf{d}(n)$, we see from (3) it is also the first column to A^\dagger , the pseudo-inverse of A . Hence, with the FTF algorithm explicitly calculating the extended gain vector, a column of a pseudo-inverse matrix is also calculated. Because the pseudo-inverse can be a discontinuous function of its perturbations, we have a cause for concern. In the next section we review theorems that pertain to this area.

3. PERTURBATION BOUNDS

Section 2 motivated attention to focus on one of the four transversal filters that together are used as the initial stage to derive the FTF algorithm for solving a least squares problem in adaptive filtering. To describe the arithmetic assumed, exact or computer, the terms method and algorithm, respectively, are used [8]. As a method, the FTF provides usable solutions, however, when implemented into a computer, its associated algorithm suffers from explosive divergence. The roundoff errors produced through computer arithmetic computations cause the theoretical relationships existing between update parameters to be lost. This in turn can understandably cause large or small perturbations to the underlying problem being solved, perturbations that can even change its rank. In short, four related problems are exploited to derive a fast algorithm whose update parameters only implicitly remain interrelated (backward consistent). Computationally, this backward consistency is not guaranteed and when lost the original four related problems that ultimately produced the algorithm become unrelated. We can therefore argue that the roundoff errors will have the effect, in a backward sense, of perturbing each of these four transversal filters. As detailed in [4], [3] and [5], this is indeed enough to demonstrate the divergence of the FTF algorithm. This reasoning falls short, however, for explaining the explosive nature of the divergence. For this we need to consider the sensitivity of the pseudo-inverse matrix of A to perturbations in the data matrix A . Although the FTF algorithm does not directly introduce perturbations into the data matrix, we can consider the computed extended gain vector $\tilde{\mathbf{k}}_M(n)$ as the exact value corresponding to the solution to (2) with a perturbed matrix \tilde{A} .

Formally, we let P and R denote the projection matrices onto the column and row space of A and $\tilde{A} = A + \delta A$, δA the matrix of perturbations associated with A having as projection matrices \tilde{P} and \tilde{R} . Then the following definition [9, pg. 139] distinguishes relatively harmless perturbations from those which can cause the pseudo-inverse matrix to behave discontinuously as a function of perturbations made to the original matrix.

Definition 1 *The matrix \tilde{A} is an acute perturbation of A if $\|\tilde{P} - P\|_2 < 1$ and $\|\tilde{R} - R\|_2 < 1$.*

An equivalent criteria for \tilde{A} and A to be acute is for $\text{rank}(A) = \text{rank}(\tilde{A}) = \text{rank}(P\tilde{A})$.

In the next subsections, we will consider the effect on the least squares solution $\mathbf{k}_M(n)$ in (3) when A is perturbed. For the special right hand side vector given in (1), we must consider the effect of these perturbations when computing A^\dagger .

3.1. Acute Perturbations

In this subsection we will assume that \tilde{A} and A are acute. For perturbations δA small in norm with respect to $\|A^\dagger\|_2$, the following assures us that the perturbed pseudo-inverse cannot suffer from explosive growth.

Theorem 1

If $\text{rank}(\tilde{A}^\dagger) = \text{rank}(A^\dagger) = r$, and $\eta = \|A^\dagger\|_2 \|\delta A\|_2 < 1$, then

$$\|\tilde{A}^\dagger\|_2 \leq \frac{1}{1-\eta} \|A^\dagger\|_2.$$

For the proof we refer to Björck [10, pg. 26]. The problems of interest require that A be of full rank, i.e., $\text{rank}(A) = M = \min((n - M + 1), M)$. Under this assumption we find that $\lim_{\delta A \rightarrow 0} \tilde{A}^\dagger = A^\dagger$. This follows from the following theorem.

Theorem 2 If $\text{rank}(\tilde{A}) = \text{rank}(A)$, then

$$\|\tilde{A}^\dagger - A^\dagger\|_2 \leq \sqrt{2} \|\tilde{A}^\dagger\|_2 \|A^\dagger\|_2 \|\delta A\|_2.$$

For the proof we refer to Wedin [11]. We note that the result holds if, in addition to an acute perturbation, $\text{rank}(A)$ is less than M . The only modification needed is to replace the coefficient $\sqrt{2}$ with $(1 + \sqrt{5})/2$.

3.2. Perturbations that are not acute

The following result is due to Wedin [11].

Theorem 3 If \tilde{A} and A are not acute, then

$$\|\tilde{A}^\dagger - A^\dagger\|_2 \geq \frac{1}{\|\delta A\|_2}$$

For $\text{rank}(\tilde{A}) \geq \text{rank}(A)$, then

$$\|\tilde{A}^\dagger\|_2 \geq \frac{1}{\|\delta A\|_2}.$$

As discussed by Stewart and Sun in [9, pg. 140], if the perturbation δA is sufficient to induce a change in rank when considering A , then A^\dagger can be described as a point of discontinuity or, in some ways, even a pole. In the context of perturbations made to A that are from round-off errors and the resultant breakdown to the relationships between parameters (backward consistency breakdown), Theorem 3 tells us that when the extended gain vector is perturbed in such a way as to have a significant component in the direction of the singular vector associated to the smallest singular value of \tilde{A} , any attempt of the FTF to reduce the perturbation will necessarily cause $\mathbf{k}_M(n)$ to diverge explosively.

4. AN EXACT SOLUTION GIVEN BY $\tilde{\mathbf{K}}_M(N)$

Since the FTF iteratively computes $\tilde{\mathbf{k}}_M(n)$ with associated residual $\tilde{\mathbf{r}}_M(n) = \mathbf{1}_{(n-M+1,1)}(n) - A\tilde{\mathbf{k}}_M(n)$ rather than the minimal residual $\mathbf{r}_M(n) = \mathbf{1}_{(n-M+1,1)}(n) - A\mathbf{k}_M(n)$, we might ask for the nearest perturbed problem (i.e., the perturbation $\|\delta A\|_2$ is a minimum) such that $\tilde{\mathbf{k}}_M(n) = \min_{\mathbf{k}_M(n)} (\mathbf{1}_{((n-M+1),1)}(n) - \tilde{A}\mathbf{k}_M(n))$.

When the least squares problem (3) is consistent (i.e., the right hand side $\mathbf{1}_{(n-M+1,1)}(n)$ is in the column span of A) we have from Regal and Gaches [12] the rank one perturbation $\delta \tilde{A}_{\min}$,

$$\begin{aligned} \delta \tilde{A}_{\min} &= \tilde{\mathbf{r}}_M(n) \tilde{\mathbf{k}}_M(n)^t / \|\tilde{\mathbf{k}}_M(n)\|_2 \\ &= \tilde{\mathbf{r}}_M(n) \tilde{\mathbf{k}}_M(n)^\dagger. \end{aligned} \quad (4)$$

This result tells us that with respect to the 2-norm, the nearest least squares problem to (3) that the computed (extended) gain vector solves exactly is in norm equal to

$$\begin{aligned} \|\delta \tilde{A}_{\min}\|_2 &= \|\tilde{\mathbf{r}}_M(n)\|_2 \|\tilde{\mathbf{k}}_M(n)^\dagger\|_2 \\ &= \frac{\|\tilde{\mathbf{r}}_M(n)\|_2}{\|\tilde{\mathbf{k}}_M(n)\|_2}. \end{aligned} \quad (5)$$

In general,

$$\|\delta \tilde{A}\|_2 = \|\tilde{\mathbf{r}}_M(n)\|_2 \|\tilde{\mathbf{k}}_M(n)\|_2$$

is called the normwise backward error.

From (5) it is possible to get small and large perturbations $\|\delta \tilde{A}_{\min}\|_2$. For example, with a small residual (with respect to the computed extended gain vector), the resultant perturbation to A can indeed be small. This is just one possibility. Of importance, however, is 1) when $\tilde{\mathbf{k}}_M(n)$ relates to \tilde{A} with rank different than A , and 2) $\|\delta \tilde{A}\|_2$ is small. With Theorem 3 it is then easy to see how $\|\mathbf{k}_M(n) - \tilde{\mathbf{k}}_M(n)\|_2$ can become large during a single update.

5. THE CONVERSION FACTOR $\gamma_M(N)$

Earlier, we mentioned that Lin [1] along with Cioffi and Kailath [2] reported that the conversion factor $\gamma_M(n)$ exceeded its theoretical bound just prior to explosive divergence. Considering the first row to (2) we have

$$1 - \mathbf{u}_M(n)^t \mathbf{k}_M(n).$$

It is this residual that is termed the conversion factor $\gamma_M(n)$ and formally is given as

$$\gamma_M(n) = 1 - \mathbf{u}_M(n)^t \mathbf{k}_M(n).$$

With this we can combine the effects from $\tilde{\mathbf{k}}_M(n)$, an associated perturbation δA to A that is not acute and Theorem 3 to describe a scenario explaining why the conversion factor could be used as a predictor for imminent explosive divergence of the FTF algorithm.

6. ALGORITHMS EXPLICITLY UPDATING

$$\mathbf{K}_M(N)$$

Given that the FTF algorithm has been modified to the so-called *stabilized* FTF algorithm [5], [13] and the problem of explosive divergence has been apparently solved (however, to date a proof of its numerical stability is lacking), these results are nonetheless important since any algorithm that updates the extended gain vector is based upon a problem that has a potential for catastrophe. If the algorithm generates a computed solution in which the associated, or inherent perturbation associated with (2) is not acute, there is potential for explosive divergence. The fact that this does not occur only means that the perturbations associated with (2) remain acute.

7. CONCLUSION

Due to the problem that the FTF algorithm has regarding consistency, that is, the updated parameters satisfy certain relationships in exact arithmetic, and these relationships are broken when the computations are performed in computer arithmetic, the FTF algorithm would be expected to diverge. However, given Theorem 1 as well as equations (4) and (5), the divergence would not be expected to be explosive. To explain this explosive divergence, we considered the well known interpretation of the extended gain vector $\mathbf{k}_M(n)$ as the solution to a special least squares problem that revealed this vector as a column to the pseudo-inverse matrix associated with the data matrix A .

Using established perturbation theory for the pseudo-inverse matrix we were able to explain why explosive divergence can occur. Additionally, through the connection between the conversion factor parameter $\gamma_M(n)$ and the extended gain vector $\mathbf{k}_M(n)$, we showed why this parameter could be more sensitive to forthcoming explosive divergence and even used as a predictor to avoid it.

8. REFERENCES

- [1] D. W. Lin, "On digital implementation of the fast Kalman algorithm," *IEEE Trans. Acoust. Speech Signal Process.*, vol. ASSP-32, pp. 998–1005, 1984.
- [2] J. M. Cioffi and T. Kailath, "Fast, recursive least squares transversal filters for adaptive filtering," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. ASSP-32, pp. 304–37, 1984.
- [3] D. T. M. Slock, "Backward consistency concept and round-off error propagation dynamics in recursive least-squares algorithms," *Opt. Engr.*, vol. 31, no. 6, pp. 1153–1169, June 1992.
- [4] P. Regalia, "Numerical stability issues in fast least-squares adaptation algorithms," *Opt. Engr.*, vol. 31, no. 6, pp. 1144–1152, June 1992.
- [5] D. T. M. Slock and T. Kailath, "Numerically stable fast transversal filters for recursive least squares adaptive filtering," *IEEE Trans. Signal Process.*, vol. 39, pp. 92–114, 1991.
- [6] J. R. Bunch, R. C. Le Borne, and I. K. Proudler, "Measuring and maintaining consistency," *Int. Journal of Applied Math and Comp. Sci.*, vol. 11, no. 5, 2001.
- [7] S. Haykin, *Adaptive Filter Theory*, Prentice-Hall, Englewood Cliffs, NJ, 2nd edition, 1991.
- [8] J. R. Bunch, R. C. Le Borne, and I. K. Proudler, "A conceptual framework for consistency, conditioning and stability issues in signal processing," *IEEE Transactions on Signal Processing*, vol. 49, no. 9, pp. 1971–1981, 2001.
- [9] G. W. Stewart and Ji-guang Sun, *Matrix Perturbation Theory*, Academic Press, 1990.
- [10] Björck, Åke, *Numerical Methods for Least Squares Problems*, Siam, Philadelphia, PA, 1996.
- [11] P. Å. Wedin, "Perturbation theory for pseudo-inverses," *BIT*, vol. 13, pp. 217–232, 1973.
- [12] J. L. Rigal and J. Gaches, "On the compatability of a given solution with the data of a linear system," *J. Assoc. Comput. Mach.*, vol. 14, pp. 543–548, 1967.
- [13] D. T. M. Slock and T. Kailath, "Numerically stable fast transversal filters for recursive least squares adaptive filtering," *IEEE Trans. Signal Processing*, vol. SP-39, pp. 92–114, 1991.