Vector Fitting vs. Levenberg-Marquardt : Some experiments

Conference Paper · June 2009

DOI: 10.1109/SPI.2009.5089858 · Source: IEEE Xplore

3 authors:

Luc Knockaert
Ghent University
196 PUBLICATIONS **1,944** CITATIONS

SEE PROFILE

Francesco Ferranti
Institut Mines-Télécom
136 PUBLICATIONS **851** CITATIONS

SEE PROFILE

Tom Dhaene
Ghent University
476 PUBLICATIONS **6,519** CITATIONS

SEE PROFILE

VECTOR FITTING VS. LEVENBERG-MARQUARDT : SOME EXPERIMENTS

Luc Knockaert[†], *Senior Member, IEEE,* Francesco Ferranti, and Tom Dhaene, *Senior Member,IEEE.*
IBBT-INTEC, Ghent University, Gaston Crommenlaan 8, PB 201, B-9050 Gent, Belgium

Abstract

A discussion and evaluation of the Vector Fitting technique is presented, in comparison with a special flavor of the Levenberg-Marquardt method, which is one of the most successful quasi Gauss-Newton techniques. Although generally slower than Vector Fitting, the Levenberg-Marquardt method can achieve greater accuracy than Vector Fitting, because of its unique ability to navigate between local minima.

1 INTRODUCTION

It is well-known that the iterative macromodeling technique known as Vector Fitting (VF) [1], is basically a reformulation of the Sanathanan-Koerner (SK) iteration [2] using a partial fraction basis [3]. The robustness of the method is mainly due to the use of partial fractions instead of polynomials, which is numerically advantageous if the initial poles are properly chosen. The VF technique has been widely used in the power systems and microwave engineering scientific communities, although it has not been as well accepted in the system identification community. The reason for this is that the mathematical problem of convergence of SK (how fast and to what?) is not very well understood theoretically [4]. Therefore, the system identification community largely prefers mainstream mathematical techniques such as Gauss-Newton (GN) [5] or quasi Gauss-Newton methods. In this contribution we discuss and evaluate the VF technique and compare it with a special flavor of the Levenberg-Marquardt (LM) method [6, 7], which is one of the most successful quasi Gauss-Newton techniques. Applied to two canonical rational approximation examples, the results are quite interesting. Although generally slower than VF, the LM method can achieve greater accuracy than VF, because of its special ability to navigate between local minima. In addition, adding an LM tail to initial VF iterations can sometimes yield a large improvement, but also sometimes not, since it is not a priori sure that good initial or final values for VF are good initial values for LM.

2 RATIONAL MODELING

Starting with M frequency data points $\{\omega_i, F_i\}$ and a suitable weighted Euclidian norm $\|\cdot\|$ over the data, the problem of proper rational modeling can be defined as the non-linear numerator-denominator optimization problem

$$\min_{N,D} \left\| F - \frac{N}{D} \right\|^2 \tag{1}$$

where the numerator $N(s)$ is a polynomial of exact degree n and the denominator $D(s)$ is a monic (coefficient of the leading term s^n is 1) polynomial of exact degree n in the Laplace variable s. Starting with an initial denominator D_0 we obtain the well-known SK [2] iterative scheme

$$\min_{N,D} \left\| \frac{FD - N}{D_0} \right\|^2 \tag{2}$$

which at each updating step $D_0 \Leftarrow D$ is a weighted linear least squares problem. The initial (generally bad) choice $D_0 = 1$ yields what is known as Levi's [8] estimate. It should be noted that the iteration idea present in SK is in fact somewhat older and due to Loeb [9]. To simplify the notation we put

$$D_0(s) = \prod_{k=1}^{n} (s - p_k) \tag{3}$$

with supposedly different poles $\{p_k\}$, and define

$$A(s) = \frac{D(s)}{D_0(s)} = a_0 + \sum_{k=1}^{n} \frac{a_k}{s - p_k} \tag{4}$$

$$B(s) = \frac{N(s)}{D_0(s)} = b_0 + \sum_{k=1}^{n} \frac{b_k}{s - p_k} \tag{5}$$

with $a_0 = 1$ ($D(s)$ is monic). Note that the representation (4-5) is always possible, since it is equivalent with the barycentric interpolation [10] formula :

$$\frac{N(s)}{D(s)} = \frac{B(s)}{A(s)} = R(s) = \frac{w_0 R(\infty) + \sum_{k=1}^{n} \frac{w_k\, R(p_k)}{s - p_k}}{w_0 + \sum_{k=1}^{n} \frac{w_k}{s - p_k}} \tag{6}$$

which correctly interpolates $R(s)$ at the poles $\{p_k\}$ for all possible weights $w_k \neq 0$. Equations (4-5) imply (see also [3]) that the SK scheme can be written as the simple Levi-like or Kalman equation error [11] linear least squares problem

$$\min_{a_i, b_i} \|FA - B\|^2 \tag{7}$$

which must iteratively be 'pole relocated', i.e., after the coefficients a_k and b_k are obtained from (7), a new pole basis $\{\tilde{p}_k\}$ is found by rooting $A(s)$, e.g. via an eigenvalue problem, and new basis functions are obtained by replacing $\{p_k\}$ with $\{\tilde{p}_k\}$ in (4-5) and so on. The rationale for pole relocation resides principally in the partial fraction expansion for rational functions with simple poles, i.e.,

$$\frac{b_0 + \sum_{k=1}^{n} \frac{b_k}{s - p_k}}{a_0 + \sum_{k=1}^{n} \frac{a_k}{s - p_k}} = \frac{\tilde{b}_0 + \sum_{k=1}^{n} \frac{\tilde{b}_k}{s - \tilde{p}_k}}{1} \tag{8}$$

so the B/A description with (old) pole basis $\{p_k\}$ corresponds with the special case $\tilde{B}/1$ for the (new) pole basis $\{\tilde{p}_k\}$. The

[†]This work was supported by a grant of the Research Foundation-Flanders (FWO-Vlaanderen).

pole relocation technique is the most salient feature of the popular VF [1] approach, in that the user can actually see, if the iterations converge, the (basis) poles settle down in their final positions. Of course, the SK technique is known, even if it converges, not to solve the original problem (1), but a related Brouwer fixed-point problem [4]. A theoretically better known technique is the GN method [5] , which is very similar to the Whitfield estimator (WE) [12]. In the GN-WE approach we linearize N/D, yielding

$$\frac{N}{D} = \frac{N_0 + \delta N}{D_0 + \delta D} \tag{9}$$

$$\approx \frac{N_0}{D_0} + \frac{\delta N}{D_0} - \frac{N_0}{D_0^2}\,\delta D \tag{10}$$

$$= \frac{N_0}{D_0} + \frac{N}{D_0} - \frac{N_0}{D_0^2}\,D \tag{11}$$

Equation (10) is the incremental formula for use in the GN method, while equation (11) is the non-incremental formula (since $N = N_0 + \delta N$ and $D = D_0 + \delta D$) for use in the WE method. Putting $F_0 = N_0/D_0, A = D/D_0, B = N/D_0$, the WE iterations with pole relocation can be written as

$$\min_{a_i,b_i} \|F_0 A - B + F - F_0\|^2 \tag{12}$$

While (7) and (12) are seemingly very different, they are not that different, since

$$FA - B = F(s) + F(s)\sum_{k=1}^{n}\frac{a_k}{s-p_k} - b_0 - \sum_{k=1}^{n}\frac{b_k}{s-p_k} \tag{13}$$

and

$$F_0 A - B + F - F_0 = F(s) + F_0(s)\sum_{k=1}^{n}\frac{a_k}{s-p_k} - b_0 - \sum_{k=1}^{n}\frac{b_k}{s-p_k} \tag{14}$$

The only difference between (13) and (14) is the presence of the rational function $F_0(s)$, which is the estimate of $F(s)$ pertaining to the previous iteration. This closeness between SK and WE was also revealed in [13].

3 THE LEVENBERG-MARQUARD METHOD

The Levenberg-Marquardt method [6, 7] is a quasi Gauss-Newton method with adaptive regularization for least squares problems. Suppose the functional to be minimized is

$$G(\mathbf{x}) = \|\mathbf{g}(\mathbf{x})\|^2 \tag{15}$$

Linearization of $\mathbf{g}(\mathbf{x})$ yields

$$\mathbf{g}(\mathbf{x} + \delta\mathbf{x}) \approx \mathbf{g}(\mathbf{x}) + \mathcal{J}_x\mathbf{h} \tag{16}$$

where $\mathcal{J}_x$ is the Jacobian of $\mathbf{g}(\mathbf{x})$ calculated at $\mathbf{x}$ and $\mathbf{h} = \delta\mathbf{x}$. The plain Gauss-Newton method finds the search step $\mathbf{h}$ by minimizing

$$\mathcal{L}_x(\mathbf{h}) = \|\mathbf{g}(\mathbf{x}) + \mathcal{J}_x\mathbf{h}\|^2 \tag{17}$$

while Levenberg-Marquardt optimization [14, 15] consist of minimizing the regularized Tikhonov functional

$$\mathcal{T}_x(\mathbf{h}) = \|\mathbf{g}(\mathbf{x}) + \mathcal{J}_x\mathbf{h}\|^2 + \mu\|\mathbf{h}\|^2 \tag{18}$$

and afterwards updating the positive damping parameter μ.

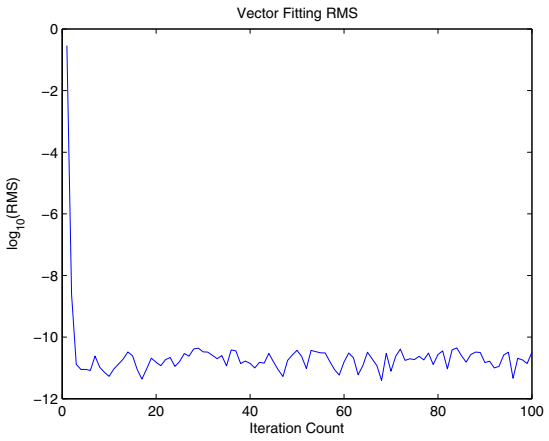

Figure 1: Vector Fitting RMS for the first example.

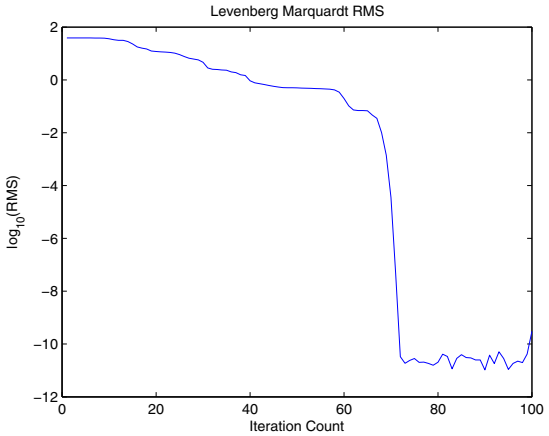

Figure 2: Levenberg Marquardt RMS for the first example.

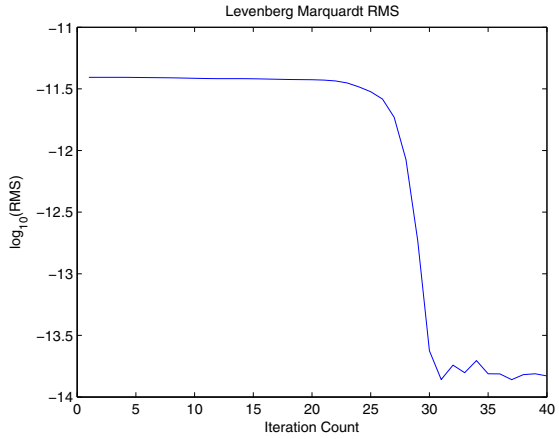

Figure 3: Levenberg Marquardt RMS for the first example with the optimal VF pole basis.

Note that practical and robust algorithms for minimizing the Tikhonov functional (18) can be found in [16]. An important quantity in the updating strategy for μ is the gain factor

$$\rho_x = \frac{G(\mathbf{x}) - G(\mathbf{x}+\mathbf{h})}{G(\mathbf{x}) - \mathcal{L}_x(\mathbf{h})} \tag{19}$$

A much used updating strategy (we call this strategy $\mathcal{S}_1$) is

$$\begin{array}{llll} \text{if} & \rho_x < \rho_1 & \text{then} & \mu \Leftarrow \beta \cdot \mu \\ \text{if} & \rho_x > \rho_2 & \text{then} & \mu \Leftarrow \mu / \gamma \\ \text{if} & \rho_x > 0 & \text{then} & \mathbf{x} \Leftarrow \mathbf{x} + \mathbf{h} \end{array}$$

where $0 < \rho_1 < \rho_2 < 1$. The BOOST parameter β and DROP parameter γ [15] must always exceed unity. Popular choices are $\rho_1 = 0.25$, $\rho_2 = 0.75$, $\beta = 2$ and $2 \leq \gamma \leq 10$. A smoother updating strategy (strategy $\mathcal{S}_2$) was proposed in [14] :

$$\begin{array}{l} \text{if} \quad \rho_x > 0 \quad \text{then} \\ \quad \mathbf{x} \Leftarrow \mathbf{x} + \mathbf{h} \\ \quad \mu \Leftarrow \mu \cdot \max\left\{1/\gamma, 1 - (\beta - 1)(2\rho_x - 1)^p\right\} \\ \quad \nu \Leftarrow \beta \\ \text{else} \\ \quad \mu \Leftarrow \mu \cdot \nu \\ \quad \nu \Leftarrow 2 \cdot \nu \end{array}$$

with ν initialized to β and p is an odd integer. In the examples we will always take $p = 3$. Strategy $\mathcal{S}_2$ also guarantees that $1/\gamma \leq \mu_{\text{new}}/\mu_{\text{old}} \leq \beta$, but without the jumps across ρ_1 and ρ_2 which are present in strategy $\mathcal{S}_1$.
In the rational approximation with pole basis we have

$$G(\mathbf{x}) = \left\| F - \frac{B}{A} \right\|^2 \qquad \mathbf{x} = [b_0, \cdots, b_n, a_0, \cdots a_n] \tag{20}$$

$$\mathcal{L}_x(\mathbf{h}) = \|F - F_0 - \frac{\delta B}{A_0} + F_0 \frac{\delta A}{A_0}\|^2 \tag{21}$$

and of course

$$\delta A(s) = \delta a_0 + \sum_{k=1}^{n} \frac{\delta a_k}{s - p_k} \tag{22}$$

$$\delta B(s) = \delta b_0 + \sum_{k=1}^{n} \frac{\delta b_k}{s - p_k} \tag{23}$$

Note that we work with a fixed pole basis in the LM method, i.e., no intermediate pole relocation steps are performed.

4 NUMERICAL SIMULATIONS

In our experiments, we found that the Levenberg-Marquardt strategy $\mathcal{S}_1$ was very slow or even stalled, while the smoother strategy $\mathcal{S}_2$ (with $\beta = 2$ and $\gamma \in [3, 8]$) always worked well. Hence we will always use strategy $\mathcal{S}_2$ in our simulations.
Our first simulation deals with the classic 18 pole example in [1] (p. 1053, Sec. 4), but without the linear term Es. The frequency range is 100 MHz and the number of equispaced sample points is 128. The approximation order is exact, i.e., we fit a rational model with 18 poles on the data. It is seen from Fig. 1 that VF (with relaxed non-triviality constraint for the improvement of convergence performance [17]) RMS settles down to its (approximate) minimal value after less than 10 iterations and afterwards behaves rather erratically. The minimal Vector Fitting $\log_{10}$(RMS) is -11.40.The time to perform the 100 VF iterations was 2.21 s. For the Levenberg Marquardt method with the initial fixed pole basis, DROP parameter $\gamma = 3$ and initial damping $\mu_0 = 1$ (a poor choice of μ_0 is not overly critical [15]), it is seen from Fig. 2 that LM RMS settles down to its (approximate) minimal value after about 70 iterations and afterwards behaves more or less erratically. The minimal Levenberg Marquardt $\log_{10}$(RMS) is -10.98. The time to perform the 100 LM iterations was 2.02 s. As a last experiment we performed LM starting with the optimal values and pole basis generated by VF, i.e., the VF pole basis resulting in the optimal VF RMS. The result is seen in Fig. 3 : the $\log_{10}$(RMS) starts with the optimal VF value -11.40, and after remaining quite static in the first 25 iterations, plunges to the new minimum -13.86 at the 31'st iteration. This means a more than hundredfold reduction of the optimal VF RMS.

As a second example we discuss the rational approximation

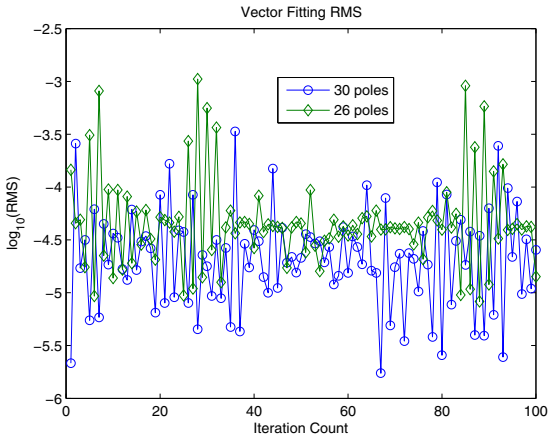

Figure 4: Vector Fitting RMS for the second example.

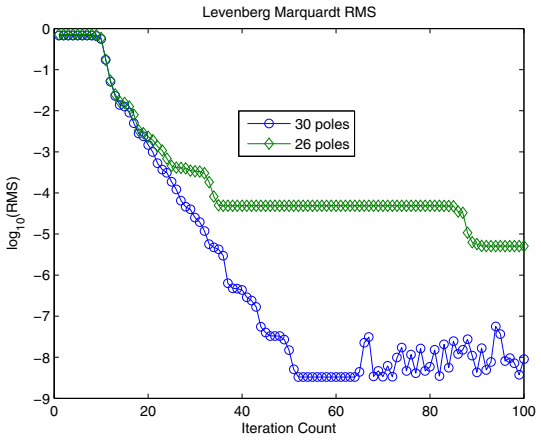

Figure 5: Levenberg Marquardt RMS for the second example.

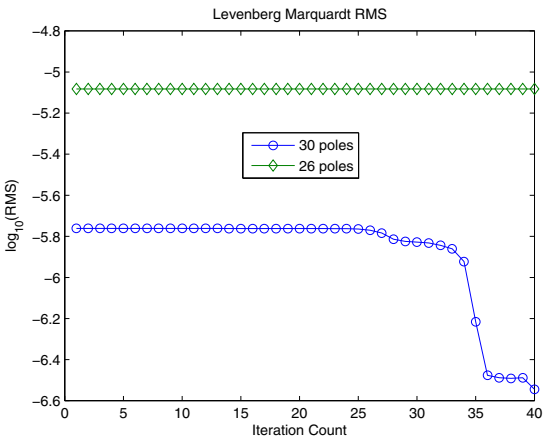

Figure 6: Levenberg Marquardt RMS for the second example with the optimal VF pole basis.

of a pure delay (transfer function $e^{-s\tau}$) with delay $\tau = 10\,\mu$s. The frequency range is 1 GHz and the number of equispaced sample points is 256. The approximation order is 30, i.e., we fit a rational model of 30 poles on the data. To indicate how the approximation order might influence the RMS, we also added the simulations for 26 poles. It is seen from Fig. 4 that VF (again with relaxed non-triviality constraint) RMS settles down to its (approximate) minimal value after one iteration (although the exact minimum is obtained only at the 67'th iteration) and afterwards behaves rather erratically. The minimal Vector Fitting $\log_{10}$(RMS) is -5.76.The time to perform the 100 VF iterations was 7.09 s. For the Levenberg Marquardt method with $\mu_0 = 0.1$ and $\gamma = 8$, it is seen from Fig. 5 that LM RMS settles smoothly down to its minimal value after 53 iterations and afterwards stalls. The minimal Levenberg Marquardt $\log_{10}$(RMS) is -8.48. The time to perform the 100 LM iterations was 9.83 s. As in the first example, we also performed LM starting with the VF pole basis resulting in the optimal VF RMS. The result is shown in Fig. 6 : it is seen that in this case, starting with the optimal VF parameters yields a small improvement for the 30 pole case, and no improvement for the 26 pole case.

5 CONCLUSION

We discussed and evaluated the Vector Fitting technique and compared it with a special flavor of the Levenberg-Marquardt method. Although generally slower than Vector Fitting, the Levenberg-Marquardt method achieved more accuracy than Vector Fitting, because of its special ability to navigate between local minima. In addition, adding a Levenberg-Marquardt tail to initial Vector Fitting iterations can sometimes yield a large improvement, but not always, since it is not a priori sure that good initial or final values for Vector Fitting are good initial values for Levenberg-Marquardt. The overall score however, taken over several iterations, may be in favor of Levenberg-Marquardt, since it has a much more smooth and non-erratic convergence behavior.

References

1. B. Gustavsen and A. Semlyen, "Rational approximation of frequency domain responses by Vector Fitting", *IEEE Trans. Power Delivery,* vol. 14, no. 3, pp. 1052-1061, July 1999.
2. C. K. Sanathanan and J. Koerner, "Transfer function synthesis as a ratio of two complex polynomials," *IEEE Trans. Automat. Contr.,* vol. 8, no. 1, pp. 56-58, Jan. 1963.
3. W. Hendrickx and T. Dhaene, "A discussion of 'Rational approximation of frequency domain responses by Vector Fitting'," *IEEE Trans. Power Systems,* vol. 21, no. 1, pp. 441-443, Feb. 2006.
4. P. A. Regalia, M. A. M. Mboup, " On the existence of stationary points for the Steiglitz-McBride algorithm," *IEEE Trans. Automat. Control,* vol. 42, no. 11, pp. 1592-1596, 1997.
5. D. S. Bayard, "High-order multivariable transfer function curve fitting: algorithms, sparse matrix methods and experimental results," *Automatica,* vol. 30, no. 9, pp. 1439-1444, 1994.
6. K. Levenberg, "A method for the solution of certain nonlinear problems in least squares," *Quart. Appl. Math.,* vol. 2, pp. 164-168, 1944.
7. D. W. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," *J. Soc. Indust. Appl. Math.,* vol. 11, pp. 431-441, 1963.
8. E. C. Levi, "Complex Curve Fitting," *IRE Trans. Automat. Contr.,* vol. AC-4, pp. 37-44, 1959.
9. H. L. Loeb, "A note on rational function approximation," *Convair Astronautics Applied Mathematics Series,* No. 27, Sept. 1959.
10. L. Knockaert, "A simple and accurate algorithm for barycentric rational interpolation," *IEEE Signal Processing Letters,* vol. 15, pp. 154-157, 2008.
11. R. E. Kalman, "Design of a self-optimizing control system," *Trans. ASME,* vol. 80, pp. 468-478, Feb. 1958.
12. A. H. Whitfield, "Asymptotic bchaviour of transfer function synthesis methods," *Int. J. Control,* vol. 45, pp. 1083-1092, 1987.
13. D. Deschrijver, T. Dhaene, G. Antonini, "A convergence analysis of iterative macromodeling methods using Whitfield's estimator," *Proc. IEEE Workshop on Signal Propagation on Interconnects, 9-12 May 2006,* pp. 197 - 200.
14. H.B. Nielsen, "Damping parameter in Marquardt's method." Report IMM-REP-1999-05, IMM, DTU. Available at `http://www2.imm.dtu.dk/documents/ftp/tr99/tr05_99.pdf`
15. M. Lampton, "Damping-undamping strategies for the Levenberg-Marquardt nonlinear least-squares method," *Computers in Physics,* vol. 11, no. 1, pp. 110-115, Jan./Feb. 1997.
16. P. C. Hansen, "Regularization tools: a Matlab package for analysis and solution of discrete ill-posed problems," *Numer. Algorithms,* vol. 6, no. 1-2, pp. 1-35, 1994.
17. B. Gustavsen, "Improving the pole relocating properties of vector fitting," *IEEE Trans. Power Delivery,* vol. 21, no. 3, pp. 1587-1592, July 2006.