Convergence Characteristics of the Variable Projection Method for Mode Extraction

Adam B. Birchfield, *Student Member, IEEE,* Thomas J. Overbye, *Fellow, IEEE*
Department of Electrical and Computer Engineering
University of Illinois at Urbana-Champaign
Urbana, IL, USA
birchfi2@illinois.edu, overbye@illinois.edu

***Abstract*—This paper reviews the variable projection method (VPM) for power system modal analysis and analyzes the method's convergence characteristics. The purpose of the VPM and other modal analysis tools is to decompose time series data into damped or undamped sinusoidal components, which provide insights into the dynamics of a measured or simulated disturbance. The paper gives five example cases of modal analysis with VPM, varying in size from a single synthetic signal to a 30-signal, 7-mode data set from simulations of a large actual power system. The analysis provides an initial indication that the VPM often finds a solution close to the matrix pencil initialization, and that the method's convergence speed can vary. While the inner loop of the method, the line search, is robust and quick, the outer VPM loop sometimes converges linearly or slower, requiring hundreds of iterations. Simpler cases with fewer modes tend to have a more consistent convergence, and are less sensitive to the initial modes selected.**

Keywords—Power system dynamics, modal analysis, convergence, power system transient stability, variable projection method

I. Introduction

Extracting the modal components of power system signals can provide insights into the dynamic characteristic and status of an electric grid. Signal data for modal analysis can come from actual measurements such as synchrophasors and digital fault recorders, or from transient stability simulations during research and planning studies. Common power system quantities these signals represent include bus voltage and frequency, branch power flow, and any of several generator control measurements. Modal analysis takes time series data from one or more of these signals, usually in the few seconds following a disturbance event, and reports a set of oscillatory components that approximately comprise the signal. Each component, or mode, is defined by its frequency of oscillation and the rate at which it is positively or negatively damped. The modes are sometimes associated with a generator's eigenvalues or the interarea resonance structure of the grid. Once the modes are extracted, analysis can be done to determine the magnitude and phase of each mode within the signals measured, which provides further insights. Thus power system planning and operations both benefit from modal identification, as engineers seek to ensure threatening oscillations are adequately damped.

Many methods have been proposed and used for performing modal identification in power systems, and the mathematical techniques behind them are even decades older. Prony analysis is a polynomial method that has been widely discussed and used [1]-[4], however, recent studies have pointed to its poor noise sensitivity and expressed a preference for other methods [5]. The matrix pencil method takes advantages of the more numerically robust and noise tolerant singular value decomposition (SVD) to identify the number of modes and extract their values, using the Hermitian matrix [5]-[6]. This method, well established in literature, is used for the initialization step in this paper.

Other more recent methods for modal analysis include the eigensystem realization method [7], the dynamic mode decomposition method [8], and the variable projection method (VPM) [9]. It is this latter method that is the topic of the present paper, a method based on nonlinear optimization to improve upon the solutions of other methods. This paper discusses an implementation of this method, its performance in example case studies, overall convergence properties, and sensitivity to the initial guess.

The rest of the paper is organized as follows. Section II reviews the theory of the VPM, along with commentary on implementing the method. Section III documents five case studies of various sizes, illustrating typical behavior of the method in synthetic signals, small test cases and a large real test case. Section IV provides an analysis of the convergence properties of the method, and Section V discusses the sensitivity of the method to the initial guess. Finally, Section VI concludes the paper.

II. The Variable Projection Method (VPM)

The VPM, like other modal analysis methods, begins with m input time signals, discretely measured at a time step of Δt, for n time points.

$$y_j(i),\, i \in 1 \dots n, j \in 1 \dots m \qquad (1)$$

The objective of modal analysis is to approximate these signals with a sum of q modes, where each mode has a frequency ω_k and a damping coefficient σ_k. Each signal y_j will also have a mode shape defined by the magnitude $A_{k,j}$ and phase $\phi_{k,j}$.

$$\hat{y}_j(i) = \sum_{k=1}^{q} A_{k,j}\, e^{\sigma_k \cdot \Delta t \cdot i} \cos(\omega_k \cdot \Delta t \cdot i + \phi_{k,j}) \qquad (2)$$

978-1-5090-6617-9/17/$31.00 ©2017 IEEE
 This paper was presented at the Texas Power and Energy Conference (TPEC), College Station, TX, February 2017.

Thus the problem is determining the best q modes (ω_k, σ_k) and $q \cdot m$ mode shapes $(A_{k,j}, \phi_{k,j})$ such that the $\widehat{y_j}$ estimations are as close as possible to the original y_j signals. It can be shown that given the modes, the shapes can be determined by a linear least-squares fit, using a simple pseudo-inversion process [6], [9]. Many other methods, such as the matrix pencil method, first solve for the modes and then approximate the shape with this linear method.

In contrast, one benefit of the VPM is that all the variables are solved for simultaneously, by projecting the linear mode shapes into the nonlinear modes. The VPM is cast as a nonlinear optimization problem,

$$\min_{\alpha} \sum_{j=1}^{m} \frac{1}{2} \left\| (\mathbf{I} - \mathbf{\Phi}(\alpha) \cdot \mathbf{\Phi}(\alpha)^{\dagger}) \cdot \mathbf{y} \right\|_2^2 \qquad (3)$$

where the optimization vector α of length p represents the mode frequencies and dampings, and the $n \times p$ matrix $\mathbf{\Phi}(\alpha)$ evaluates each mode at each point in time. The variable $p = 2 \cdot q - q_0$, where q_0 is the number of modes with $\omega_k = 0$. The matrix $\mathbf{y}$ is the $n \times m$ matrix giving the input data.

The VPM problem is solved using a steepest-direction gradient line search, where the method for finding the gradient $\nabla\alpha$ is given in [9]. The method begins with an initial guess for the modes α, and then iterates using a line search in the negative gradient direction, until the solution converges to an acceptable tolerance.

In implementing this method for the present paper, several design considerations must be taken into account. First, there may be some preprocessing of the signals to remove non-modal components. This is done on a case-by-case basis in this paper. For example, a frequency signal with a nominal value of 60 Hz will have a low magnitude of oscillations compared to the signal magnitude, and so removing the constant component may be advantageous. All signals in this paper are normalized by their standard deviation, so that they are compared on a consistent basis: for example, the 500 MW line flow signal should not overshadow the 1.0 per-unit voltage signal. After the scaling, both signals would have a standard deviation of 1.

Next, the initial values of α are important. For this paper, the matrix pencil method is used to initialize α. This step also determines the size of α and therefore the number of modes. Two parameters of the matrix pencil method are the pencil parameter L and the singular value threshold σ_t, as a fraction of the largest singular value. For this paper, L is given a value such that the Hermitian matrix is approximately square, and the value σ_t is set to 0.025 for most cases, except case 2, as described in the next section.

Once the initial modes are determined, decisions must be made about the optimization procedure. This paper implements a golden section search, where the maximum distance to be searched is some fraction of $\nabla\alpha$, which varies from 1 to 0.001 in this paper. Then convergence tolerances must be determined. Since the golden section search is rather fast, it is evaluated until the cost function is minimized within 10^{-12}. The outer loop, which is the actual VPM gradient calculation, is much slower and the tolerance must depend on the size of the system, as the following sections will discuss.

Benefits of the VPM, as discussed in [9], include that it optimizes over the final function fit, considering the mode shape which is usually done after the modes are locked-in. The method also provides for flexibility in both the input functions and the time step used. In This paper shows some examples of the method's use and other characteristics that may be worth considering when deciding which approach to take for mode extraction.

III. Five VPM Case Studies

This section documents five case studies which apply the VPM to various signals, extracting their modes and commenting on the optimization process. The first two examples are synthetic, with known modes; one also contains noise. The third and fourth examples are from the WSCC 9-bus test case [10], measuring various signals under a disturbance in both stable and unstable conditions. The final example is from an actual large power system case, with about 16,000 buses, showing how the VPM applies to studies on a big case with many input signals and possible modes.

A. *Case 1: Synthetic Single Signal Without Noise*

This case is the simplest of them all. A signal has 31 time steps spaced 0.1 s apart, giving a noise-free single mode with a frequency of 4 rad/s and a damping coefficient of 0.2, meaning it is positively damped. The mode has a magnitude of 1 and phase of 0 in the signal given.

No preprocessing was done on the signal. For the initial matrix pencil method, the threshold σ_t was set to 0.025, and the single mode was correctly identified, within a tolerance of 10^{-10} for each of the parameters. The first VPM iteration returned a cost function value of 2.8×10^{-18}, and the gradient had a magnitude of 7.6×10^{-10}, and so no VPM optimization steps were needed.

B. *Case 2: Synthetic Double Signal With Noise*

Two signals are created for this case, based on two modes. The first mode has a frequency of 6 rad/s and a damping of 0.5, while the other has a frequency of 0 and a damping of 3, meaning it is not an oscillatory mode. The first signal contains the first mode with magnitude 2 and phase 0, and contains the second mode with a magnitude of 10. The second signal contains the first mode with magnitude of 3 and phase of 0.4 radians, and contains the second mode with magnitude 5. Table I summarizes these modes. In addition, a uniform random time-invariant noise signal was added with a peak magnitude of 1. The time series given has a sampling frequency of 0.01 and a length of 201 time steps.

The signals were scaled by their standard deviations during the preprocessing, to ensure they have equivalent weight in the cost function. This scaling is compensated for in the results by multiplying the mode shape magnitudes by the signals' original standard deviations. With the threshold σ_t increased to 0.08 to avoid picking up numerous extra modes due to the noise, the resulting modes are as indicated in Table I, and the cost function evaluates to 2.3039.

TABLE I
CASE 2 MODES

	Mode 0	Mode 1
Frequency, actual (rad/s)	6	0
Frequency, initial	6.0512	--
Frequency, VPM final	6.0495	--
Damping, actual	0.5	3
Damping, initial	0.5192	2.9498
Damping, VPM final	0.5053	2.9857
S1 magnitude, actual	2	10
S1 magnitude, initial	2.0035	9.9371
S1 magnitude, VPM final	1.9628	10.0203
S1 phase, actual (radians)	0	--
S1 phase, initial	-0.01273	--
S1 phase, VPM final	-0.00890	--
S2 magnitude, actual	3	5
S2 magnitude, initial	3.0481	4.7244
S2 magnitude, VPM final	3.0092	4.7710
S2 phase, actual (radians)	0.4	--
S2 phase, initial	0.3210	--
S2 phase, VPM final	0.3241	--

The VPM iterations use a line search with the maximum distance to be searched is $\nabla\alpha$. For completeness, the VPM iterations were continued until the cost function changes less than 10^{-12}, which takes 35 iterations and 4.2 seconds. Probably actual applications would not require this much precision. At this step, the cost function has reduced to 1.9917, and the modes are shown in Table I. As can be seen in this table, the initial guess is relatively close to the actual values, considering the noise, and the VPM iterations, while they decreased the cost function by 14%, did not appreciably increase the proximity to the actual modes.

C. Case 3: WSCC 9-Bus Stable Frequency Disturbance

The WSCC 9-bus system [10] contains three generators. Case 3 measures the bus frequency at each generator, in the 10 seconds following a line faulting and opening. The signals are sampled at a time step of 0.05, and 201 time steps are used for the three signals. The preprocessing step subtracts 60.04 from each signal, to remove the dc offset. The standard 0.025 is used for σ_t, and the initialization picks out four modes at frequencies 1.098 Hz, 1.001 Hz, 0.180 Hz, and 0 Hz.

The line search width is slowed to $0.1\nabla\alpha$ since using the full gradient width causes the line search to diverge. The process is run to six digits of precision, which takes 770 iterations and about 4 minutes. The cost function decreases from 0.4288 to 0.3166. The change from the initial to final modes is shown in Table II. Similar to the previous example, the modes do not change dramatically. One interesting trend in this case is that the VPM increases the damping of each mode, and compensates by increasing each mode's magnitude in the various signals.

D. Case 4: WSCC 9-Bus Unstable Voltage Disturbance

The WSCC 9-bus system is used again in case 4, with the per-unit voltage magnitude measured at all 9 buses this time. The disturbance is similar to case 3, however the generator at bus 2 has its exciter feedback gain increased to make the system unstable. The 9 signals are sampled at the simulation time step of 0.008333 s, for a 10s window, which leads to 10809 total data points, on the order of the studies done in [9]. The preprocessing step subtracts 1.0 from each signal and scales each by its standard deviation. The initialization picks out four modes, one of which is negatively damped.

TABLE II
CASE 3 MODES

Mode	Initial	VPM final
0, Frequency (Hz)	1.0982	1.1077
0, Damping	0.5357	0.7816
1, Frequency	1.0088	1.0093
1, Damping	0.1937	0.2663
2, Frequency	0.1900	0.1869
2, Damping	0.4902	0.5012
3, Frequency	0	0
3, Damping	1.0034	1.2665

TABLE III
CASE 4 MODES

Mode	Initial	VPM final
0, Frequency (Hz)	0.9368	0.9462
0, Damping	0.4452	0.3818
1, Frequency	0.1768	0.1775
1, Damping	0.4481	0.4900
2, Frequency	0	0
2, Damping	0.01625	0.02413
3, Frequency	0	0
3, Damping	-0.15235	-0.14997

For this case, the line search width must be slowed again to $0.01\ \nabla\alpha$ because the search diverges in the first iteration otherwise. The process is run again to six digits of precision, and again it takes about 5 minutes to complete, but this time only 19 iterations are required. The modes can be seen in Table III, including mode 3, which shows the negatively damped, unstable portion of the signals. The modes do not tend to change dramatically, but the damping changes more than the frequency.

E. Case 5: Large System Study

This final case uses time series data from a transient stability simulation using a large actual power system test case with about 16,000 buses. The contingency is a fault and opening of a large inter-area transmission line, and thirty signals are saved for analysis: ten bus voltages, ten generator speeds, and ten branch active power flows. The sampling time step is 0.025s for a 10s ringdown period, leading to 12,030 data points. The system is stable under this contingency. For preprocessing, each signal was subjected to a linear regression, with the least-squares linear line of best fit removed from the signal, and following this the signals were scaled according to their standard deviations, so that in the cost function they would be equivalent.

The matrix pencil initialization process results in seven modes, which the VPM iterations change more significantly than in the other cases, as Table IV shows. The cost function was reduced by the VPM method from 198.64 to 30.21, a reduction of 85%. The acceleration factor for the gradient is reduced to 0.001 in this case, and the iterations were run until the change in cost function was less than 10^{-3}, which took 392 iterations and about 100 minutes. It should be noted that the cost function was reduced 77% of the 85% within 2 iterations, which could be completed in less than one minute.

TABLE IV
CASE 5 MODES

Mode	Initial	VPM final
0, Frequency (Hz)	1.1045	1.1414
0, Damping	0.9514	1.0819
1, Frequency	0.8955	0.8653
1, Damping	0.9714	1.2715
2, Frequency	0.6484	0.6520
2, Damping	0.2181	0.5308
3, Frequency	0.3642	0.3370
3, Damping	0.4424	0.2748
4, Frequency	0.3106	0.3630
4, Damping	0.5606	0.9842
5, Frequency	0.1688	0.2061
5, Damping	0.2725	0.2554
6, Frequency	0.0341	0.0361
6, Damping	0.0207	-0.0354

The greatest change in frequency for this case was mode 5, which changed 22%, while the greatest change in damping were modes 2 and 6. Mode 6 is particularly interesting, because the VPM changed a slightly damped mode at a very low frequency into one at approximately the same frequency that is very slightly undamped.

IV. CONVERGENCE ANALYSIS

The previous five case studies show example VPM applications in a variety of time steps, number of signals, and type of data. It is clear that there is a variety in the solution as well, in terms of the change in modes as well as the number of iterations and time required to reach a certain level of convergence. This section provides an initial analysis into the characteristics of these solutions and their implications about the method's performance.

A. Inner Loop Line Search

As noted in Section II, each VPM iteration solves for the gradient, and then a golden section line search is performed to find the optimal distance to travel in the negative gradient direction. Fig. 1 illustrates the space of this search with the first iteration of case 5. This figure shows a search space that is approximately quadratic in shape, with a clear minimum. This general shape is typical for a VPM iteration, which makes the convergence properties of the inner loop quite fast, provided the search window is wide enough to include the minimum and narrow enough to exclude numerical issues, which occur beyond 0.001 in the given plot. So long as this window is found, the convergence rate is quadratic, and the minimum is found within a tolerance of 10^{-12} in less than 40 iterations throughout each VPM iteration in each case. For case 3, the average number of iterations is 20.6, and for case 5, the average is 25.3. This is consistent with a quadratic convergence rate, where one additional digit of precision is obtained for every two to three iterations. Fig. 2 shows the golden section search process, where the window length decays exponentially with the number of iterations. This portion of the method seems quick and robust, provided a sufficient original window length can be found.

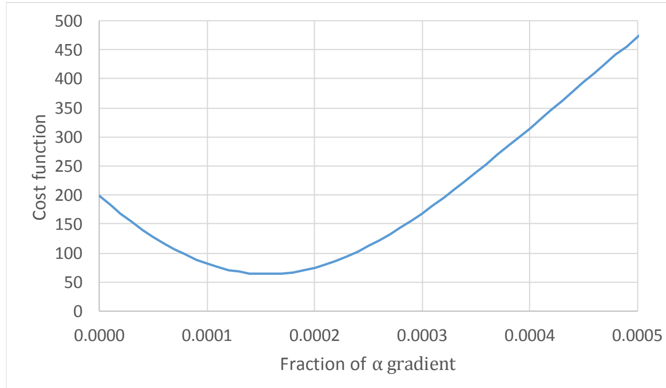

Fig. 1. First VPM iteration of case 5, showing the cost function with respect to the distance in the negative gradient direction.

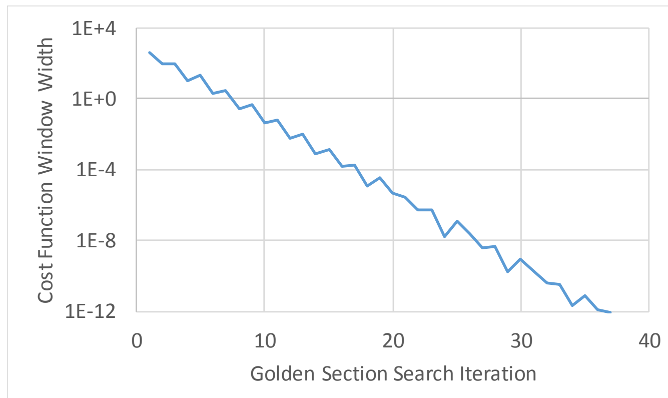

Fig. 2. First VPM iteration of case 5, showing the decrease in window width through the iterations of the golden section search (inner loop).

B. Outer Loop VPM

When it comes to the convergence of the outer VPM loop, the five cases show a variety in the response characteristics. Case 1 acutely illustrates an observation that the rest of the cases confirm: the matrix pencil initialization provides a very good guess of the modal content. In the perfect, noiseless case, the VPM convergence is immediate because the initialization can find the modes accurately. Under noise, as case 2 and others show, matrix pencil performs relatively well too, and the objective of the VPM would be to improve upon an already good solution.

Figs. 3-6 show the convergence rates of each solution for cases 2-5. These figures plot the change in cost function as the iterations progress. There is a clear difference between cases 2 and 4, which converged relatively quickly, and cases 3 and 5, which very slowly converged. In cases 3 and 5, the first few iterations showed rapid improvement, but after that the convergence was linear or worse.

The VPM follows a steepest-descent algorithm, which is known to have a convergence rate that can sometimes be very slow, since following the instantaneous steepest-descent direction can lead to zig-zagging paths, depending on the

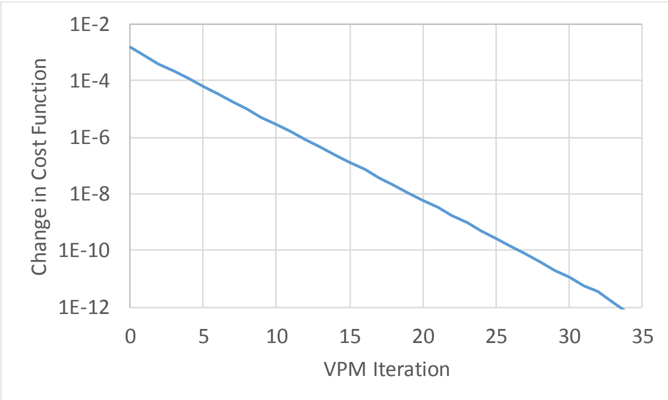

Fig. 3. Convergence of case 2.

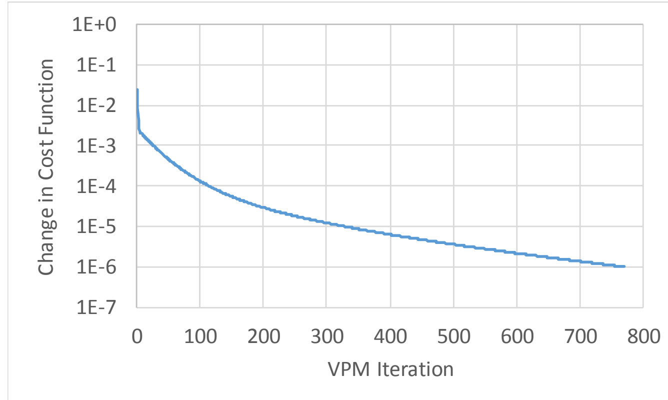

Fig. 4. Convergence of case 3.

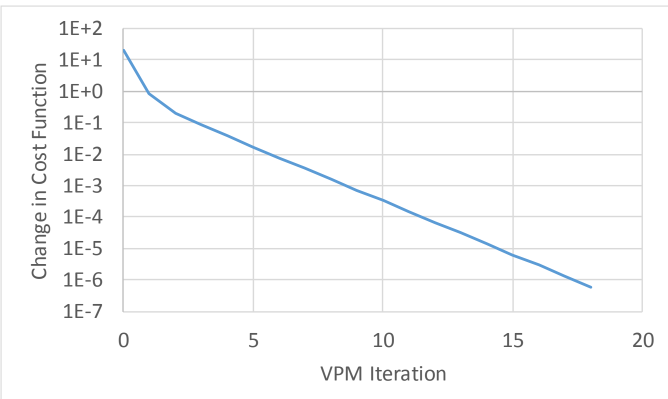

Fig. 5. Convergence of case 4.

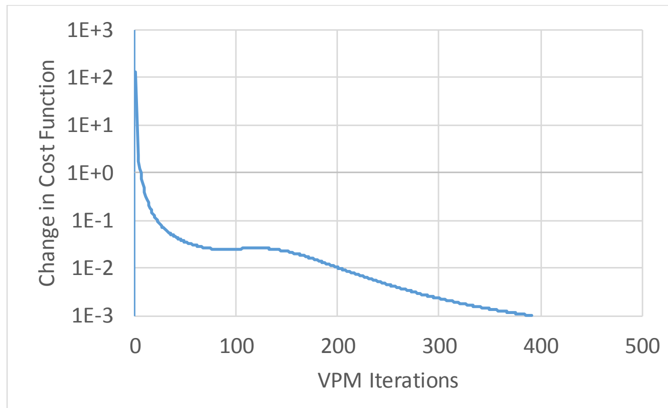

Fig. 6. Convergence of case 5.

convexity and eccentricity of the optimization gradient. The rate may change during the process as well.

It is clear that the VPM can sometimes have a quick, robust convergence rate, as in cases 2 and 4, where any logarithm of convergence tolerance can be reached with a small linear number of iterations. Other times VPM will converge very slowly, where an additional decimal of convergence will require hundreds of iterations or more. That rate may change throughout the process, as illustrated particularly by case 5. In case 5, the first few iterations show fast convergence, but the convergence stalls and even decreases from iterations 91-118, before increasing again at a much slower rate, as shown in Fig. 6.

Different factors may contribute to the convergence rates of the VPM. A small number of time points, such as in case 2, may help, although case 3 has fewer time points than case 4. The number of modes is important, as case 5 converted to quadratic convergence when the initialization threshold was changed to only allow 2 modes instead of 7. But this is not a universal principle, since cases 3 and 4 have the same number of modes. Case 5 has 7 modes, which is more than the examples given in [9], and may point to convergence issues and unreliable behavior when the number of modes becomes too large.

V. Sensitivity to Initial Guess

It is important to note that VPM does not determine the number of modes, or the number of modes which are non-oscillatory. This must be pre-determined. However, there are many options for the initial values of those modes, of which the matrix pencil method used in this paper is only one.

Case 1 serves as the first example for examining the VPM's region of convergence, because it is simple with only two variables in α. It is found that, approximately, if the initial frequency is within the range 0.01 to 20, and the damping is between -20 and 10, the VPM will eventually converge to the same, correct solution, frequency of 4 and damping of -0.2. This is a wide region of convergence, however the number of VPM iterations required to get the specified solution varies greatly. Fig. 7 illustrates this for varying initial frequencies, with the damping initialized to 0, and Fig. 8 shows the same thing for varying initial damping with an initial frequency of 3. For this case, a reasonably wide radius around the solution converges within 20 iterations.

A similar analysis is performed on case 3, with the number of convergence iterations recorded for various initialization configurations. Table V shows the results, where the initialization was equal to the VPM solution, with a slight perturbation on one variable. This process was followed for all seven variables, for perturbations of 1%, 10%, and for some variables 100%. The convergence of this case is known to be slow, so it is not surprising that most variables show slow convergence when perturbed 10%. Within a very narrow region, it appears that the convergence is quick.

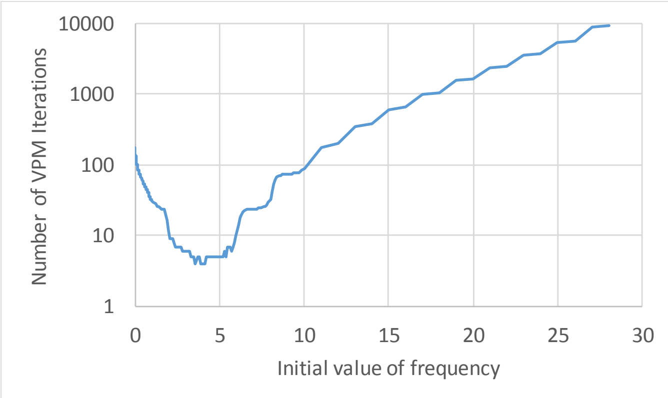

Fig. 7. Sensitivity of case 1 convergence iterations to initial guess of frequency, with the initial guess of damping set to 0.

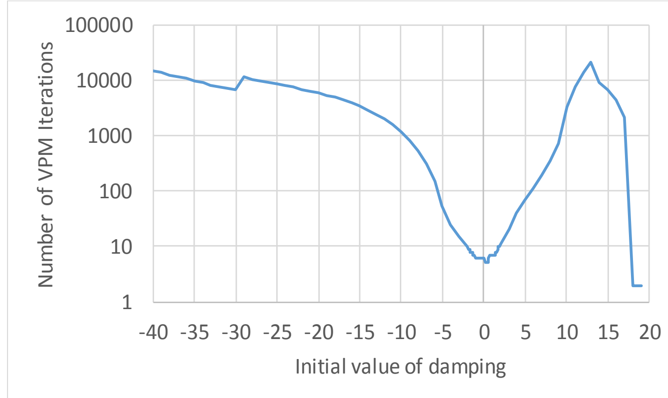

Fig. 8. Sensitivity of case 1 convergence iterations to initial guess of damping, with the initial guess of frequency set to 3. Only those initialized from -29 to 13 converged to the correct value.

TABLE V
CASE 3 MODES, NUMBER OF CONVERGENCE ITERATIONS FOLLOWING VARIOUS PERTURBATIONS

Mode	Perturbation		
	1%	10%	100%
0, Frequency (Hz)	245	1382	--
0, Damping	61	256	903
1, Frequency	12	945	--
1, Damping	7	28	296
2, Frequency	7	575	844
2, Damping	7	12	156
3, Damping	5	429	--

Cases 1 and 3 show the importance of the initialization step to obtaining an accurate solution in a reasonable amount of time. For some simple cases, as in case 1, initializing in the neighborhood of the solution will lead to quick convergence, while even initialization far from the answer is likely to converge eventually. Other cases, like case 3, converge prohibitively slowly if the initial values are not close to the final solution.

VI. CONCLUSION

The VPM is a relatively newly-developed method to be used in modal extraction for power system modal analysis in potentially both online monitoring and transient stability contingency studies during the planning process. In many cases, especially those with few modes, the method appears robust and quickly converging. Future work may improve upon the optimization procedure to provide better performance in those conditions which tend toward slower convergence.

REFERENCES

[1] J. F. Hauer, C. J. Demeure, and L. L. Scharf, "Initial results in Prony analysis of power system response signals," *IEEE Trans. Power Syst.,* vol. 5, no. 1, pp. 80-89, Feb. 1990.

[2] M. A. Johnson, I. P. Zarafonitis, and M. Calligaris, "Prony analysis and power system stability-some recent theoretical and applications research," in *Proc. 2000 IEEE Power Eng. Society Summer Meeting*, vol. 3, pp 1918-1923.

[3] J. J. Sanchez-Gasca and J. H. Chao, "Performance comparison of three identification methods for the analysis of electromechanical oscillations," *IEEE Trans. Power Syst.*, vol. 14, no. 3, pp. 995-1002, Aug. 1999.

[4] D. J. Trudnowski, J. M. Johnson, and J. F. Hauer, "Making Prony analysis more accurate using multiple signals," *IEEE Trans. Power Syst.*, vol. 14, no. 1, pp. 226-231, Feb. 1999.

[5] L. L. Grant and M. L. Crow, "Comparison of matrix pencil and Prony methods for power system modal analysis of noisy signals," in *Proc. 2011 North American Power Symposium (NAPS)*, pp. 1-7, Aug. 2011.

[6] Y. Hua and T. K. Sarkar, "Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise," in *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. 38, no. 5, pp. 814-824, May 1990.

[7] R. A. DeCallafon, B. Moaveni, J. P. Conte, X. He, and E. Udd, "General realization algorithm for modal identification of linear dynamic systems," *ASCE J. Eng. Mech.*, vol. 134, no. 9, pp. 712-722, 2009.

[8] S. Mohaptra and T. J. Overbye, "Fast modal identification, monitoring, and visualization for large-scale power systems using dynamic mode decomposition," in *Proc. 2016 Power Systems Computation Conference (PSCC),* Genoa, Italy, pp. 1-7.

[9] A. R. Borden and B. C. Lesieutre, "Variable projection method for power system modal identification," in *IEEE Trans. Power Syst.*, vol. 29, no. 6, pp. 2613-2620, Nov. 2014.

[10] http://icseg.iti.illinois.edu/wscc-9-bus-system/