Using Complex Arithmetic to Identify Groundwater Modeling Parameters

by Clarence O. E. Burg

May 20, 1999

Abstract

In groundwater modeling, the parameters that determine an aquifer's properties often change dramatically throughout the aquifer. Unfortunately, the only direct information about an aquifer's transmissivity (T) and storage coefficient (S) is available at pumping and observation wells. Due to the variability of these parameters, the values for T and S as measured at a well may not be accurate a few feet from the well and may be quite inaccurate as a measurement for the entire aquifer. However, at a well, pumping rate and water level information is also available. The water level or hydraulic head h in a well is a function of the pumping rate and of the aquifer parameters. By using this information, the distribution of parameters throughout the aquifer that yields the hydraulic head drawdown curve for a particular set of pumping rates can be estimated.

In this research, an implicit, finite-element, analysis code solves the two-dimensional, unsteady, porous media equation for a confined aquifer with a heterogeneous parameter distribution on an unstructured grid, with multiple, variable pumping rate wells, or

$$S\frac{\partial h}{\partial t} = \frac{\partial}{\partial x}\left(T\frac{\partial h}{\partial x}\right) + \frac{\partial}{\partial y}\left(T\frac{\partial h}{\partial y}\right) + \text{Pumping Terms} \tag{1}$$

A non-linear least-squares function F is formed from the measured and computed water levels at the N_w wells and for the N_t times or

$$F(T,S) = \sum_{i=1}^{N_t}\left(\sum_{j=1}^{N_w}\left(h_{i,j}^{measured} - h_{i,j}(T,S)\right)^2\right) \tag{2}$$

The derivatives of this function with respect to the parameters T and S are calculated by using the complex Taylor's series expansion method. Using this derivative information within the Gauss-Newton optimization algorithm, the parameters are updated with the goal of minimizing the function F. This method is applied to two inverse problems where the target parameter distribution is known and to one inverse problem where the target parameter distribution is unknown.

1 Overview of Groundwater Modeling

Groundwater modeling is of interest to engineers for at least two reasons. The first reason is to ensure that an aquifer is not being de-watered by the pumping wells. An aquifer is de-watered when hydraulic head in a well falls below the width of the aquifer, which means that there are regions in the aquifer where the water has been completely removed. In these regions, the ground may compress and no longer be able to hold water. Another reason for groundwater modeling is to determine the flow patterns in an aquifer in order to predict the location of contaminants released into the aquifer.

The primary difficulty in groundwater modeling is the determination of the parameters which are the transmissivity and the storage coefficient. These parameters can be determined at a well by analyzing the core sample. But they can vary widely within an aquifer. For instance, in the Gordo aquifer in Mississippi (*Slack and Darden*[1991]), tests of core samples at various wells indicate that the transmissivity varies from 130 to $24,000 ft^2/day$ (1.4 to $258 cm^2/sec$) and that the storage coefficient varies from 0.00003 to 0.006. Hence, the values of these parameters at the well are probably not a representative value for the entire aquifer.

To estimate the values of these parameters through the entire aquifer, several inverse methods have been proposed that use the drawdown curve of the hydraulic head in a well due to the pumping rates. By recording the pumping rates and the hydraulic head at a well, a large amount of data can be obtained that depend on these unknown parameters. Using this data, the parameter values that produced these drawdown curves can be approximated. *Sun and Yeh*[1992], *Kitanidis*[1997], and *Hantush and Marino*[1997] used stochastic methods, and *Townley and Wilson*[1985] used sensitivity analysis to generate the design space derivatives to solve the parameter identification problem.

In this paper, the objective function to be minimized measures the difference between the measured and the computed hydraulic heads for four pumping wells and four observation wells for several time periods. The computed heads are determined via an unstructured, finite-element simulation code that solves the two-dimensional, unsteady, heterogeneous, porous media equation, allowing for variable pumping rate wells. The derivative of the objective function is determined to machine accuracy by using the complex Taylor's series expansion method (*Squire and Trapp*[1998]). This derivative information is used in conjunction with the Gauss-Newton optimization (*Gill, Murray and Wright*[1981]) method to achieve rapid convergence.

This method is applied to two parameter identification problems where the target parameters are known and to one problem where the target parameters are not known. In the first problem, the parameter distribution is defined by values at one interior node and six exterior nodes, and the target depths are determined from this parameter distribution. In the second problem, the parameter distribution in the aquifer is defined by values at seven interior nodes and six exterior nodes, providing for better variation within the aquifer. In the third problem, the parameter distribution is defined by using the seven interior node grid but are solved using the five interior node grid; hence, the optimal values for the parameters are not known.

2 Complex Taylor's Series Expansion Method

Instead of using finite differences to approximate the derivative, the complex Taylor's series expansion method uses complex arithmetic to derive a formula that produces an estimate for the derivative that is accurate to machine accuracy. Consider the Taylor's series expansion of $F(x + i\Delta x)$ or

$$F(x + i\Delta x) = F(x) + iF'(x)\Delta x - \frac{F''(x)}{2}\Delta x^2 - i\frac{F'''(x)}{6}\Delta x^3 + O(\Delta x^4) \tag{3}$$

If $F(x)$ is a real-valued function, then the expansion easily decomposes into the real and imaginary parts. The imaginary part is

$$Im(F(x+i\Delta x)) = F'(x)\Delta x - \frac{F'''(x)}{6}\Delta x^3 + O(\Delta x^5) \tag{4}$$

By solving for $F'(x)$, we get

$$F'(x) = \frac{Im(F(x+i\Delta x))}{\Delta x} + \frac{F'''(x)}{6}\Delta x^2 + O(\Delta x^4) \approx \frac{Im(F(x+i\Delta x))}{\Delta x} \tag{5}$$

By choosing Δx such that the higher order terms are smaller than machine zero, the derivative of F can be approximated to machine accuracy. This method easily extends to functions of more than one variable.

3 Test Cases

The goal of these test cases was to determine the ability of the parameter identification method to estimate the correct parameter distribution accurately, using few iterations. For these problems, the same computational domain was used, which had four pumping wells and four observation wells. The ranges for the parameters were based on those found in the Gordo aquifer, so transmissivity varied from 0 to $1000 cm^2/sec$ and the storage coefficient varied from 0.0001 to 0.01. The computational domain was an unstructured grid, consisting of 624 nodes and 1226 elements, that allowed large changes in element size to reflect the relative sizes of the well and the aquifer. The domain was a circle with radius of 5 km and is pictured in Figure 1. The three parameter grids are shown in Figure 2 and overlay the computational domain. The values of T and S are defined at the nodes of the parameter grid, and the values in the aquifer are determined from these values via linear interpolation. The initial values of T and S were the same for each node, being set to $10^{1.5}$ and $10^{-2.0}$ respectively.

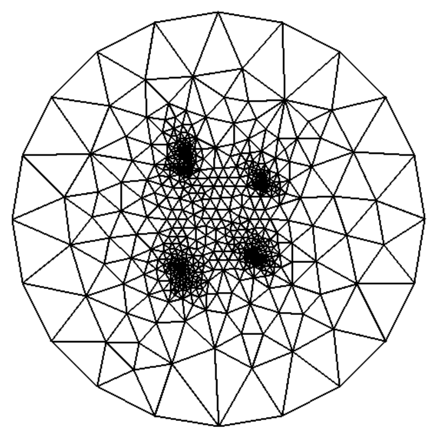

Figure 1. Computational Domain

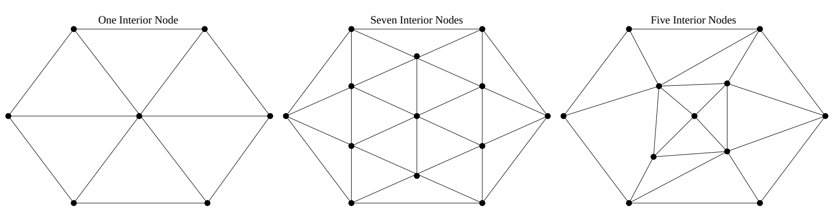

Figure 2. Parameter Distribution Grids.

For each test case, the drawdown curves for the four pumping wells and the four observation wells were determined for a particular parameter distribution grid and for a set of parameter values and stored as $h_{i,j}^{measured}$ where i refers to the time level and j refers to the well number. The objective function was defined as the least-squares difference between the measured or target values and the computed values of $h_{i,j}(T,S)$ and is given in equation (2).

3.1 Case 1 - 7 Node Parameter Grid

For this test case, the target values were determined for the parameter grid with one interior node, and the computed values were determined using the same grid. Hence, the parameter identification process should be able to identify the parameter values exactly. Once the target values were computed and stored, the identification process was started. It successfully identified the target values to machine accuracy in 10 iterations. Figure 3 shows the optimization history for this test case. The objective function was decreased by 26 orders of magnitude, demonstrating that the optimization process was quite efficient in locating the target parameter values. For iteration number 5, the objective function value was approximately 100. At this point, the values for transmissivity were accurate to approximately two or three significant digits, and the values for the storage coefficient were not quite as accurate, with some being off by 10-20%.

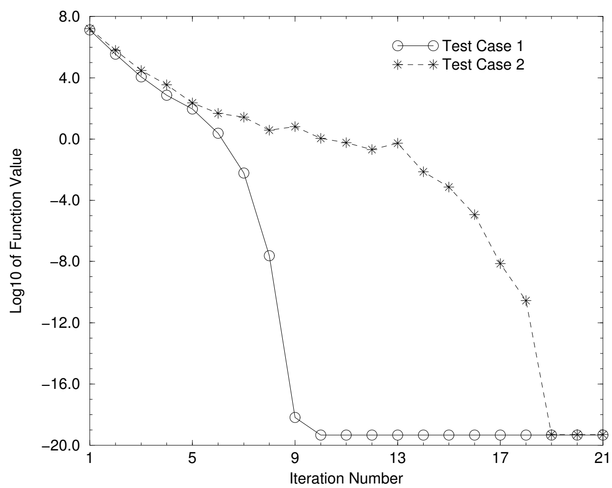

Figure 3. Optimization Histories

3.2 Case 2 - 13 Node Parameter Grid

For test case 2, the parameter distribution grid with 7 interior nodes was used to define the target values and was also used to identify the parameter values. Thus, the target values should also be identified exactly. In this test case, the transmissivity and storage coefficient were determined from parameter values defined at 13 nodes, so there was a total of 26 parameter values to identify. Thus, the parameter space was significantly larger than for test case 1, where there were only 14 parameter values, and there were more parameter space derivatives to determine. Because of the increased complexity, the parameter identification process required 19 iterations to converge to the target parameters. The optimization history for this test case is also shown in Figure 3.

3.3 Case 3 - Unknown Optimum

For the final test case, the parameter grid with 7 interior nodes was used to compute the target values, while the grid with only 5 interior nodes was used to identify the target values. As a result, the function value could not be driven to zero. However, the optimal value of the function was approximately 60, which was significantly smaller than the initial value of over 10^6. This value was on the same order of magnitude as the result discussed for Case 1, which indicated that the transmissivity was approximated to about 2 significant digits and the storage coefficient was approximated to within 10-20%.

To determine the reliability of the identified parameter values, a new set of target values was generated based on different pumping rates, which were considerably different from the previous rates. Using the new set of target values and the identified parameter values, the new function was evaluated, which was approximately twice the old value. As a result, we can conclude that the identified parameter values were relatively accurate representations of the parameter values for the entire grid.

4 Conclusions

The complex Taylor's series expansion (CTSE) method was used to convert an unstructured, two-dimensional, heterogeneous, solver of the porous media equation into an optimization code that could identify the transmissivity (T) and storage coefficient (S) based on pumping rates and drawdown curves at various wells. A non-linear least-squares function was defined as the difference between the measured and the computed drawdown curves for 8 different wells. The CTSE method was used to estimate the derivatives of this function with respect to the parameters to a high level of accuracy, and the Gauss-Newton optimization method was used to update the parameter values.

The first two test cases showed that the parameter identification method was able to identify the parameters exactly when the target values were attainable. In the third test case, the target values were not within the parameter space and hence only an approximate solution could be found. Nevertheless, the values that were found were quite accurate, as shown by repeating the problem with different pumping rates. By using more refined parameter grids, higher level of accuracy are expected.

5 References

1. Gill, P. E., Murray, W., and Wright, M. H., Practical Optimization, Academic Press, Inc., San Diego, CA, 1981.

2. Hantush, M. M., and Marino, M. A., "Stochastic Solution to Inverse Problem in Ground Water", *J. Hydr. Engr.*, Vol. 123, No. 12, Dec., 1997.

3. Kitanidis, P. K., "The Minimum Structure Solution to the Inverse Problem", *Water Res. Res.*, Vol. 33, No. 10, pp. 2263-2272, Oct. 1997.

4. Slack, L. J., and Darden, D., "Summary of Aquifer Tests in Mississippi, June 1942 through May 1988", *U.S. Geological Survey*, Water-Resources Investigations Report 90-4155, 1991.

5. Squire, W., and Trapp, G., "Using Complex Variables to Estimate Derivatives of Real Functions", *SIAM Review*, Vol. 40, No. 1, pp. 110-112, March, 1998.

6. Sun, N. Z., and Yeh, W. W.-G., "A Stochastic Inverse Solution for Transient Groundwater Flow: Parameter Identification and Reliability Analysis", *Water Res. Res.*, Vol. 28, No. 12, pp. 3269-3280, Dec., 1992.

7. Townley, L. R., and Wilson, J. L., "Computationally Efficient Algorithms for Parameter Estimation and Uncertainty Propagation in Numerical Models of Groundwater Flow", *Water Res. Res.*, Vol. 21, No. 12, pp. 1851-1860, Dec., 1985.