Solving a double porosity soil characteristic curve with the quasi-Newton method (BFGS)

St. Yannopoulos & C. Tzimopoulos

Department of Rural and Surveying Engineering, Aristotle University of Thessaloniki, Greece

Abstract

The characteristic curve of almost all the existing soils can be satisfactorily simulated by an S-shape curve. However, in cases where the medium has a double porosity (macroporous soils, unsaturated fractured rocks and fissured ground water systems), the S-shape curve fails to describe the characteristic curve. Gerke and van Genuchten [14] have described the physical structure of such a medium and Zhang and van Genuchten [33] have proposed a new model with five parameters, which can simulate the double porosity soils. In this article a special algorithm is developed in order to optimize the new function describing the double porosity characteristic curve. The algorithm is based on the theory of the quasi-Newton method (BFGS)

1 Introduction

Solution of unsaturated flow problems usually requires predetermination of the soil hydraulic properties, namely (i) the relationship between the soil pressure head, $h$, and the soil moisture content, $\theta$ and (ii) the dependence of the hydraulic conductivity, $K$, upon the soil moisture content, $\theta$. In order to determine these relationships a large number of laboratory and field methods has been developed [6], [16], [18], [19]. Common to all measurement methods is a certain conflict between accuracy of the results and the necessary expense.

A large number of laboratory and field methods have been developed over the years to measure the unsaturated hydraulic conductivity, most of them are relatively costly and difficult to implement. Specifically, accurate in-situ measurement of the unsaturated hydraulic conductivity has remained especially cumbersome and time consuming [28], [33].

As an alternative to direct measurements, theoretical methods have been developed to estimate the hydraulic properties from more easily measured soil
water retention data. These theoretical methods are generally based on statistical pore-size distribution models, which assume that water flow through cylindrical pores and incorporate the equations of Darcy and Poiseuille [28], [29].

These macroscopic prediction models estimate unsaturated hydraulic conductivity \( K \) from \( \theta \) or \( h \) based on an analogy between the micro- and the macroscale. They have a theoretical basis and generally, contain only a small number of adjustable parameters. The most important shortcomings of most macroscopic models are: (i) The neglect of the influence of a variable pore-size distribution on the hydraulic conductivity [9]. (ii) The assumption of having a spherical pore geometry [3]. However, statistical prediction models have the advantages that they take into account information about the pore-size distribution of a soil and they predict unsaturated hydraulic conductivity from the more easily measured soil moisture characteristic curve and a factor allowing for a variable pore tortuosity. A unified presentation of the statistical models is given by Mualem [21].

Implementation of these predictive conductivity models still requires independently measured soil water retention data. In the literature, while a large number of analytical soil water functions are available, only a few of them can be easily incorporated into the predictive pore-size distribution models to yield relatively simple analytical expressions for the unsaturated hydraulic conductivity, as for example equations of Brooks and Corey [3], Brutsaert [7], Campbell [8], Tzimopoulos and Sakellariou [27], van Genuchten [29], Visser [30].

As it is pointed by Zurmühl and Durner [34], one of the drawbacks of most descriptions of hydraulic functions is that these curves imply the soil to possess a unimodal distribution of equivalent pore size, i.e. the derivative of the water content, \( \theta \), with respect to the logarithm of soil pressure head, \( \log(h) \), is a function that has only one maximum. The soil pores are classified in micropores where capillarity exists and macropores without capillarity. The micropores are detectable by the soil water characteristic curve, while the macropores not. The derivative curve of the soil water characteristic curve reflects the pore-size distribution. When the micropore-size distribution has two, or generally \( n \), peaks, we are speaking about bi- or n-model porosity [22]. The macropores form a dual porosity system, where the microporous system exists next to the system of macropores [2].

Implicitly, the natural soils do not have a unimodal pore-size distribution. Soils with multimodal pore-size distribution include undisturbed soils and clays with bimodal or multimodal particle size distribution [22], [23], [29], morainic soils [11], soils well-developed secondary pore systems stemming from aggregation, and soils with a micropore system derived from roots and/or fanal action [1], [26].

The problem of the multimodal pore-size distribution study has caused the interest because of the bimodal nature of many soil pore systems and because of the importance of bimodal hydraulic functions in predicting preferential flow of water and chemicals in undisturbed soils or fracture rocks [14], [24].
As showed by Durner [11] and Othmer et al. [22], the multimodal retention functions can be described by summing sigmoidal retention curves given by van Genuchten's [29] expressions, i.e.

\[ S_{e_i} = \frac{\theta_i - \theta_{r,i}}{\theta_{s,i} - \theta_{r,i}} = \frac{1}{\left[ 1 + (\alpha_i h)^{n_i} \right]^{1-1/n_i}}, \quad i=1,2 \tag{1} \]

where \( \theta_r \) and \( \theta_s \) are the residual and saturated water contents, respectively, \( \alpha_i \) and \( n_i \) are empirical constants affecting the shape of the retention curves. The multiporosity soil moisture retention curve (1) keeps some of the desired properties of the van Genuchten [29] curve i.e. it is continuously differentiable, strongly monotonous function with zero slopes at saturation and toward the dry end. Disadvantages of the multi-porosity curve (1) are the increased number of model parameters, lack of a physical basis of the model parameters and absence of a closed-form invertible function [11]. Nevertheless, eqn (1) can describe soil moisture retention curve satisfactorily in undisturbed soils with distinct double peaks or with asymmetric pore-size distribution curves.

To describe soil moisture characteristic curve \( \theta(h) \) in a double porosity porous media, Zang and van Genuchten [33] have proposed a relatively simple nonlinear model. In order to utilize this model as input for numerical simulation of water flow and solute transport, its parameters must be estimated from observed retention data in the field or in the laboratory. In this paper a special optimization algorithm is presented for the estimation of the parameters of Zang and van Genuchten [33] model based on the theory of the quasi-Newton method (BFGS).

2 The bimodal model

Zhang and Genuchten [33] have proposed the following equation for the description of water retention data of soils exhibiting bimodal pore size distribution, which can be expressed in dimensionless form as

\[ S_e = \frac{1 + c_1 h^*}{1 + h^* + c_2 h^{*2}}, \tag{2} \]

where the reduced pressure head, \( h^* \), is given by

\[ h^* = \alpha h, \tag{3} \]

while the reduced water content (or effective saturation), \( S_e \), is defined as

\[ S_e = \frac{\theta - \theta_r}{\theta_s - \theta_r}. \tag{4} \]

In eqns (2)-(4) \( \theta \) is the water content [cm\(^3\)/cm\(^3\)], \( \theta_r \) and \( \theta_s \) are the residual and saturated water contents [cm\(^3\)/cm\(^3\)], respectively, \( \alpha \) is a scaling factor [cm\(^{-1}\), \( \alpha > 0 \)], \( h \) is the soil water pressure head [cm] assumed to be positive in this paper (suction), and \( c_1 \) and \( c_2 \) are empirical parameters affecting the shape of the
retention curve. In eqn (2), $0 \leq c_1 \leq 1$ and $c_2 > 0$ are. From eqn (2), it is evident that at saturation, where $h^* = 0$ is $Se = 1$, while for very dry conditions, where $h^* \rightarrow \infty$ the reduced water content, $Se \rightarrow 0$.

The soil moisture characteristic curve of eqn (2) changes from an S-shape curve to a increasingly bimodal curve, when $c_2$ decreases.

Inverting eqn (2), the soil moisture characteristic curve, $h(Se)$, yields two roots, from which only the positive root

$$h^* = \frac{(c_1 - Se) + \sqrt{(c_1 - Se)^2 + 4c_2 Se(1 - Se)}}{2c_2 Se}$$

ensures the conditions:

$h^* = 0$ for $Se = 1$, $h^* > 0$ for $0 < Se < 1$ and $h^* \rightarrow \infty$ for $Se = 0$.

3 Parameter estimation

Inspection of eqn (2) shows that the soil moisture characteristic curve, $B(h)$, contains five unknown parameters ($\theta$, $\theta_\ast$, $\alpha$, $c_1$ and $c_2$), which must be estimated from the measured points ($\theta_i$, $h_i$). From these parameters, the saturated water content, $\theta_\ast$, can be measured. Regarding the so-called residual water content, $\theta_r$, its physical meaning is not clear in the literature. An extended review of the estimation methods of $\theta$ was presented by Yannopoulos [32]. Supposing that $\theta_r$ is a well defined parameter and that there are sufficiently accurate estimations of $\theta_r$ and $\theta_\ast$, then the remaining three parameters $\alpha$, $c_1$ and $c_2$ can be estimated from the minimization of the function:

$$f = \sum_i (\theta_i - \theta_{i,\text{exp}})^2$$

where

$$\theta_i = \theta_r + (\theta_\ast - \theta_r) \frac{1 + c_1 \alpha h_i}{1 + \alpha h_i + c_2 (\alpha h_i)^2}$$

and $\theta_{i,\text{exp}}$ are the measured water contents.

4 Quasi-Newton Method (BFGS)

The problem of the nonlinear optimization using quasi-Newton methods can be stated as:

$$\min f(\tilde{x})$$

in which $\tilde{x}$ is a vector on $n$ decision variables $\tilde{x} = (x_1, x_2, \ldots, x_n)$.

In general search algorithms have been introduced, which directly attack the above problem and the minimization procedure consists of two steps:

- Determine the search direction along which the objective function value decreases.
Find a new point \( \bar{x}_{k+1} = \bar{x}_k + \beta_k \bar{d}_k \), where \( \bar{d}_k \) is the vector indicating the search direction and \( \beta_k \) is a scalar representing the step size, whose optimal value is to be determined. This step is called line search and mathematically the optimization can be stated as:

\[
\min f(\bar{x}_k + \beta_k \bar{d}_k).
\]

Convergence criteria or stop rules have to be introduced for determination:

\[
\| \bar{x}_{k+1} - \bar{x}_k \| < \varepsilon_1, \text{ or } \| f(\bar{x}_{k+1}) - f(\bar{x}_k) \| < \varepsilon_2 \text{ etc.}
\]

In the general Newton’s method the recursive equation for a line search is:

\[
\bar{x}_{k+1} = \bar{x}_k - H(\bar{x}_k)^{-1} \nabla f(\bar{x}_k),
\]

where \( H(\bar{x}_k) = \nabla (\nabla f(\bar{x}_k)) \) is the Hessian matrix, assumed strictly positive defined.

The main disadvantage of the Newton’s method is that it requires the inversion of the Hessian matrix in each iteration, which is a cumbersome task. For this reason Davidon [10] and later Fletcher and Powell [12] proposed the quasi-Newton method.

The search directions are of the form:

\[
\bar{S}_k = -n(\bar{x}_k)\nabla f(\bar{x}_k) \text{ in lieu of } -H(\bar{x}_k)^{-1} \nabla f(\bar{x}_k)
\]

as in Newton’s method. The gradient direction is thus deflected by premultiplying it by \(-n(\bar{x}_k)\), where \(n(\bar{x}_k)\) is a \(nxn\) positive definite symmetric matrix that approximates the inverse of the Hessian matrix.

So in each iteration one proceeds to construct the new matrix as follows:

\[
n(\bar{x}_{k+1}) = n(\bar{x}_k) + \Delta n(\bar{x}_k)
\]

where

\[
\Delta n(\bar{x}_k) = \frac{\bar{P}_k \bar{P}_k^T}{\bar{P}_k^T \bar{q}_k} \frac{n(\bar{x}_k)q_k \cdot q_k^T}{q_k^T \cdot n(\bar{x}_k) \cdot q_k}.
\]

The vectors \(\bar{P}_k\) and \(\bar{q}_k\) are defined as:

\[
\bar{P}_k = \bar{x}_{k+1} - \bar{x}_k, \quad \bar{q}_k = \nabla f(\bar{x}_{k+1}) - \nabla f(\bar{x}_k).
\]

A useful generalization of the Davidon-Fletcher-Powell method was proposed by Broyden [4]. Essentially, Broyden [4] introduced a degree of freedom in updating the matrix \(n(\bar{x}_k)\). A particular choice of this degree of freedom was proposed independently by Broyden [5], Fletcher [13], Goldfarb [15] and Shano [25]. This updating, known as the BFGS update has been consistently shown in
many computational studies to dominate other updating schemes in its overall performance. The updating correction $\Delta n(\bar{x}_k)$ is:

$$\Delta n(\bar{x}_k) = \left[1 + \frac{\tilde{q}_k^T n(\bar{x}_k) \tilde{q}_k}{\tilde{P}_k^T \tilde{q}_k} \right] \cdot \frac{\tilde{P}_k^T \tilde{q}_k}{\tilde{P}_k^T \tilde{q}_k} - \left[\tilde{P}_k^T \tilde{q}_k n(\bar{x}_k) + n(\bar{x}_k) \tilde{q}_k \tilde{P}_k^T \right] \frac{\tilde{P}_k^T \tilde{q}_k}{\tilde{P}_k^T \tilde{q}_k}.$$

5 Results and discussions

Four soils of the literature, of different hydraulic properties, were subjected to the analysis of this paper. The data for three of them, namely Touchet Silt Loam, Silt Mont Cenis and Beit Netofa Clay were taken from Mualem’s catalogue [20]. The data for the remaining one soil, namely Sarpy Loam was taken from Hanks and Bowers [17].

Table 1: Values of parameters

<table>
<thead>
<tr>
<th>Soil</th>
<th>$\theta_i$</th>
<th>$\theta_r$</th>
<th>$a$</th>
<th>$c_1$</th>
<th>$c_2$</th>
<th>$\text{mse}$</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sarpy Loam</td>
<td>0.410</td>
<td>0.060</td>
<td>0.0316</td>
<td>0.842</td>
<td>0.3595</td>
<td>$7 \times 10^{-3}$</td>
<td>0.997</td>
</tr>
<tr>
<td>Touchet Silt Loam</td>
<td>0.520</td>
<td>0.000</td>
<td>0.0014</td>
<td>0.010</td>
<td>200</td>
<td>$2.56 \times 10^{-4}$</td>
<td>0.980</td>
</tr>
<tr>
<td>Silt Mont Cenis</td>
<td>0.442</td>
<td>0.017</td>
<td>0.0060</td>
<td>0.392</td>
<td>0.020</td>
<td>$3.09 \times 10^{-5}$</td>
<td>0.998</td>
</tr>
<tr>
<td>Beit Netofa Clay</td>
<td>0.490</td>
<td>0.000</td>
<td>0.017</td>
<td>0.900</td>
<td>0.012</td>
<td>$1.33 \times 10^{-5}$</td>
<td>0.99</td>
</tr>
</tbody>
</table>

Figure 1: Observed and fitted retention curves, $\theta(h)$, for Sarpy Loam
The values of parameters $\alpha$, $c_1$ and $c_2$ obtained from the present optimization algorithm are shown in Table 1. Also, in the same table the $\theta_0$ and $\theta_r$ -values, the mean square error, mse, and the correlation coefficient, $R^2$, are given.

Based on the values of the parameters $\alpha$, $c_1$ and $c_2$ obtained from the present algorithm, curves $\theta(h)$ are computed and plotted against experimental points in figures 1-4. In all these figures it can be observed the very good agreement between measured and calculated retention curves.

Figure 2: Observed and fitted retention curves, $\theta(h)$, for Touchet Silt Loam

Figure 3: Observed and fitted retention curves, $\theta(h)$, for Silt Mont Cenis
From the results of figures 1 and 3, we can see the excellent agreement of the calculated retention curves with the present algorithm and with the conjugate directions one (Yannopoulos and Tzimopoulos, [31]).

Figure 4: Observed and fitted retention curves, \( \theta(h) \), for Beit Netofa Clay

References


Water Resources Management


