CHAPTER 1

Introduction to Unsaturated Soil Mechanics

Soil mechanics involves a combination of engineering mechanics and the properties of soils. This description is broad and can encompass a wide range of soil types. These soils could either be saturated with water or have other fluids in the voids (e.g., air). The development of classical soil mechanics has led to an emphasis on particular types of soils. The common soil types are saturated sands, silts and clays, and dry sands. These materials have been the emphasis in soil mechanics textbooks. More and more, it is realized that attention must be given to a broader spectrum of materials. This can be illustrated by the increasing number of research conferences directed towards special classes of soil types and problems.

There are numerous materials encountered in engineering practice whose behavior is not consistent with the principles and concepts of classical, saturated soil mechanics. Commonly, it is the presence of more than two phases that results in a material that is difficult to deal with in engineering practice. Soils that are unsaturated form the largest category of materials which do not adhere in behavior to classical, saturated soil mechanics.

The general field of soil mechanics can be subdivided into that portion dealing with saturated soils and that portion dealing with unsaturated soils (Fig. 1.1). The differentiation between saturated and unsaturated soils becomes necessary due to basic differences in their nature and engineering behavior. An unsaturated soil has more than two phases, and the pore-water pressure is negative relative to the pore-air pressure. Any soil near the ground surface, present in a relatively dry environment, will be subjected to negative pore-water pressures and possible desaturation.

The process of excavating, remolding, and recompacting a soil also results in an unsaturated material. These materials form a large category of soils that have been difficult to consider within the framework of classical soil mechanics.

Natural surficial deposits of soil are at relatively low water contents over a large area of the earth. Highly plastic clays subjected to a changing environment have produced the category of materials known as swelling soils. The shrinkage of materials may pose an equally severe situation. Loose silty soils often undergo collapse when subjected to wetting, and possibly a loading environment. The pore-water pressure in both of the above cases is initially negative, and volume changes occur as a result of increases in the pore-water pressure.

Residual soils have been of particular concern in recent years. Once again, the primary factor contributing to their unusual behavior is their negative pore-water pressures. Attempts have been made to use saturated soil mechanics design procedures on these soils with limited success.

An unsaturated soil is commonly defined as having three phases, namely, 1) solids, 2) water, and 3) air. However, it may be more correct to recognize the existence of a fourth phase, namely, that of the air-water interface or contractile skin (Fredlund and Morgenstern, 1977). The justification and need for a fourth phase is discussed later in this chapter. The presence of even the smallest amount of air renders a soil unsaturated. A small amount of air, likely occurring as occluded air bubbles, renders the pore fluid compressible. Generally, it is a larger amount of air which makes the air phase continuous throughout the soil. At the same time, the pore-air and pore-water pressures begin to differ significantly, with the result that the principles and concepts involved differ from those of classical, saturated soil mechanics. These differing conditions are addressed throughout this book.

1.1 ROLE OF CLIMATE

Climate plays an important role in whether a soil is saturated or unsaturated. Water is removed from the soil either by evaporation from the ground surface or by evaportranspiration from a vegetative cover (Fig. 1.2). These processes produce an upward flux of water out of the soil. On the other hand, rainfall and other forms of precipitation provide a downward flux into the soil. The difference be-
between these two flux conditions on a local scale largely dictates the pore-water pressure conditions in the soil.

A net upward flux produces a gradual drying, cracking, and desiccation of the soil mass, whereas a net downward flux eventually saturates a soil mass. The depth of the water table is influenced, amongst other things, by the net surface flux. A hydrostatic line relative to the groundwater table represents an equilibrium condition where there is no flux at ground surface. During dry periods, the pore-water pressures become more negative than those represented by the hydrostatic line. The opposite condition occurs during wet periods.

Grasses, trees, and other plants growing on the ground surface dry the soil by applying a tension to the pore-water through evapotranspiration (Dorsey, 1940). Most plants are capable of applying 1–2 MPa (10–20 atm) of tension to the pore-water prior to reaching their wilting point (Taylor and Ashcroft, 1972). Evapotranspiration also results in the consolidation and desaturation of the soil mass.

The tension applied to the pore-water acts in all directions, and can readily exceed the lateral confining pressure in the soil. When this happens, a secondary mode of desaturation commences (i.e., cracking).

Year after year, the deposit is subjected to varying and changing environmental conditions. These produce changes in the pore-water pressure distribution, which in turn result in shrinking and swelling of the soil deposit. The pore-water pressure distribution with depth can take on a wide variety of shapes as a result of environmental changes (Fig. 1.2).

Significant areas of the earth's surface are classified as arid zones. The annual evaporation from the ground surface in these regions exceeds the annual precipitation. Figure 1.3 shows the climatic classification of the extremely arid, arid, and semi-arid areas of the world. Meigs (1953) used the Thornthwaite moisture index (Thornthwaite, 1948) to map these zones. He excluded the cold deserts. Regions with a Thornthwaite moisture index less than −40 indicate arid areas. About 33% of the earth's surface is considered arid and semi-arid (Dregne, 1976). The distribution of extremely arid, arid, and semi-arid areas in North America is shown in Fig. 1.4. These areas cover much of the region bounded by the Gulf of Mexico in the south, up into Canada in the north, over to the west coast.

Arid and semi-arid areas usually have a deep groundwater table. Soils located above the water table have neg-

---

**Figure 1.1** Categorization of soil mechanics.

**Figure 1.2** Stress distribution during the desiccation of a soil.
1.2 TYPES OF PROBLEMS

The types of problems of interest in unsaturated soil mechanics are similar to those of interest in saturated soil mechanics. Common to all unsaturated soil situations are the negative pressures in the pore-water. The type of problem involving negative pore-water pressures that has received the most attention is that of swelling or expansive clays. However, an attempt is made in this book to broaden the scope of problems to which the principles and concepts of unsaturated soil mechanics can be applied.

Several typical problems are described to illustrate relevant questions which might be asked by the geotechnical engineer. Throughout this book, an attempt is made to respond to these questions, mainly from a theoretical standpoint.

1.2.1 Construction and Operation of a Dam

Let us consider the construction of a homogeneous rolled earth dam. The construction involves compacting soil in approximately 150 mm (6 in) lifts from its base to the full height of the dam. The compacted soil would have an initial degree of saturation of about 80%. Figure 1.5 shows a dam at approximately one half of its design height, with a lift of soil having just been placed. The pore-air pressure in the layer of soil being compacted is approximately equal to the atmospheric pressure. The pore-water pressure is negative, often considerably lower than zero absolute pressure.

The soil at lower elevations in the fill is compressed by the placement of the overlying fill. Each layer of fill constitutes an increase in total stress to the embankment. Compression results in a change in the pore-air and pore-water pressures. The construction of the fill is generally rapid enough that the soil undergoes volume change under essentially undrained conditions. At any time during construction, the pore-air and pore-water pressures can be contoured as shown in Fig. 1.6.

In reality, some dissipation of the pore pressures will occur as the fill is being placed. The pore-air pressure will
dissipate to the atmosphere. The pore-water pressure may also be influenced by evaporation and infiltration at the surface of the dam. All pore pressure changes produce volume changes since the stress state is being changed.

There are many questions that can be asked, and there are many analyses that would be useful to the geotechnical engineer. During the early stages of construction, some relevant questions are:

- What is the magnitude of the pore-air and pore-water pressure induced as each layer of fill is placed?
- Is pore-air pressure of significance?
- Does the engineer only need to be concerned with the pore-water pressures?
- Does an induced pore-air pressure result in an increase or a decrease in the stability of the dam? Or would the computed factor of safety be conservative if the pore-air pressures are assumed to be zero?
- What is the effect of air going into solution and subsequently coming out of solution?
- Will the pore-air pressure dissipate to atmospheric conditions much faster than the pore-water pressures can come to equilibrium?
- What deformations would be anticipated as a result of changes in the total stress and the dissipation of the induced pore-air and pore-water pressures?
- What are the boundary conditions for the air and water phases during the placement of the fill?

Once the construction of the dam is complete, the filling of the reservoir will change the pore pressures in a manner similar to that shown in Fig. 1.7. This indicates a transient process with new boundary conditions. Some questions that might be asked are:

- What are the boundary conditions associated with the equalization processes once the filling of the reservoir is underway?
- How will the pore-air and pore-water pressures change with time, and what are the new equilibrium conditions?
- Will further deformation take place as the pore-air and pore-water pressures change in the absence of a change in total stress? If so, how much deformation can be anticipated as steady-state conditions are established?
1.2 TYPES OF PROBLEMS

Numbers are pore-water pressures (kPa)

Figure 1.7 Typical pore-water and pore-air pressures after some dissipation of pore pressures and partial filling of the reservoir.

- What changes take place in the limit equilibrium factor of safety of the dam as the reservoir is being filled and pore-water pressures tend to a steady-state condition?

After steady-state conditions are established, changes in the environment may give rise to further questions (Fig. 1.8).

- Does water flow across the phreatic surface under steady-state conditions?
- What effect will a prolonged dry or wet period have on the pore pressures in the dam?
- Could a prolonged dry period produce cracking of the dam? If so, to what depth might the cracks extend?
- Could a prolonged wet period result in the local or overall instability of the dam?

Answers to all of the above questions involve an understanding of the behavior of unsaturated soils. The questions involve analyses associated with saturated/unsaturated seepage, the change in volume of the soil mass, and the change in shear strength. The change in the shear strength could be expressed as a change in the factor of safety. These questions are similar to those asked when dealing with saturated soils; however, there is one primary difference. In the case of unsaturated soils problems, the flux boundary conditions produced by changes in the environment play a more important role.

1.2.2 Natural Slopes Subjected to Environmental Changes

Natural slopes are subjected to a continuously changing environment (Fig. 1.9). An engineer may be asked to investigate the present stability of a slope, and predict what would happen if the geometry of the slope were changed or if the environmental conditions should happen to change. In this case, boreholes may be drilled and undisturbed samples obtained for laboratory tests. Most or all of the potential slip surfaces may lie above the groundwater table. In other words, the potential slip surface may pass through unsaturated soils with negative pore-water pressures. Typical questions that might need to be addressed are:

- What effect could changes in the geometry have on the pore pressure conditions?

Figure 1.8 The effect of rainfall on steady-state flow through a dam.
INTRODUCTION TO UNSATURATED SOIL MECHANICS

Soil stratum 1

Soil stratum 2

Soil stratum 3

Phreatic surface

Figure 1.9 An example of the effect of excavations on a natural slope subjected to environmental changes.

- What changes in pore pressures would result from a prolonged period of precipitation? How could reasonable pore pressures be predicted?
- Could the location of a potential slip surface change as a result of precipitation?
- How significantly would a slope stability analysis be affected if negative pore-water pressures were ignored?
- What would be the limit equilibrium factor of safety of the slope as a function of time?
- What lateral deformations might be anticipated as a result of changes in pore pressures?

Similar questions might be of concern with respect to relatively flat slopes. Surface sloughing commonly occurs on slopes following prolonged periods of precipitation. These failures have received little attention from an analytical standpoint. One of the main difficulties appears to have been associated with the assessment of pore-water pressures in the zone above the groundwater table.

The slow, gradual, downslope creep of soil is another aspect which has not received much attention in the literature. It has been observed, however, that the movements occur in response to seasonal, environment changes. Wetting and drying, freezing and thawing are known to be important factors. It would appear that an understanding of unsaturated soil behavior is imperative in formulating an analytical solution to these problems.

1.2.3 Mounding Below Waste Retention Ponds

Waste materials from mining and industry operations are often stored as a liquid or slurry retained by low-level dikes (Fig. 1.10). Soil profiles with a deep water table are considered to be ideal locations for these waste ponds. The soils above the water table have negative pore-water pressures and may be unsaturated. It has often been assumed that as long as the pore-water pressure remained negative, there is little or no movement of fluids downward from the waste pond. However, in recent years, it has been observed that a mounding of the water table may occur below the waste pond, even though the intermediate soil may remain unsaturated. Now, engineers realize that significant volumes of water and contaminants can move through the soil profile, even though negative pore-water pressures are retained.

Questions of importance with respect to this type of problem would be:

- How should seepage be modeled for this situation? What are the boundary conditions?
- How should the coefficient of permeability of the unsaturated soil be characterized? The coefficient of permeability is a function of the negative pore-water pressure, and thereby becomes a variable in a seepage analysis.
- What equipment and procedures should be used to characterize the coefficient of permeability in the laboratory?
- How do the contaminant transport numerical models interface with unsaturated flow modeling?
- What would be the effect on the water table mounding if a clay liner were placed at the base of the retention pond?

1.2.4 Stability of Vertical or Near Vertical Excavations

Vertical or near vertical excavations are often used for the installation of a foundation or a pipeline (Fig. 1.11). It is well known that the backslope in a moist silty or clayey soil will stand at a near vertical slope for some time before failing. Failure of the backslope is a function of the soil type, the depth of the excavation, the depth of tension cracks, the amount of precipitation, as well other factors. In the event that the contractor should leave the excavation open longer than planned or, should a high precipitation period be encountered, the backslope may fail, causing damage and possible loss of life.

The excavations being referred to are in soils above the groundwater table where the pore-water pressures are negative. The excavation of soil also produces a further decrease in the pore-water pressures. This results in an increase in the shear strength of the soil. With time, there
1.2 TYPES OF PROBLEMS

Precipitation Potential

Figure 1.11 An example of potential instability of a near vertical excavation during the construction of a foundation.

will generally be a gradual increase in the pore–water pressures in the backslope, and correspondingly, a loss in strength. The increase in the pore–water pressure is the primary factor contributing to the instability of the excavation. Engineers often place the responsibility for ensuring backslope stability onto the contractor. Predictions associated with this problem require an understanding of unsaturated soil behavior.

Some relevant questions that might be asked are:
- How long will the excavation backslope stand prior to failing?
- How could the excavation backslope be analytically modeled, and what would be the boundary conditions?
- What soil parameters are required for the above modeling?
- What in situ measurements could be taken to indicate incipient instability? Also, could soil suction measurements be of value?
- What effect would a ground surface covering (e.g., plastic sheeting) have on the stability of the backslope?
- What would be the effect of temporary bracing, and how much bracing would be required to ensure stability?

1.2.5 Lateral Earth Pressures

Figure 1.12 shows two situations where an understanding of lateral earth pressures is necessary. Another situation might involve lateral pressure against a grade beam placed on piles. Let us assume that in each situation, a relatively dry clayey soil has been placed and compacted. With time, water may seep into the soil, causing it to expand in both a vertical and horizontal direction. Although these situations may illustrate the development of high lateral earth pressures, they are not necessarily good design procedures.

Some questions that might be asked are:
- How high might the lateral pressures be against a vertical wall upon wetting of the backfill?
- What are the magnitudes of the active and passive earth pressures for an unsaturated soil?
- Are the lateral pressures related to the “swelling pressure” of the soil?

Figure 1.12 Examples of lateral earth pressures generated subsequent to backfilling with dry soils. (a) Lateral earth pressures against a retaining wall as water infiltrates the compacted backfill; (b) lateral earth pressure against a house basement wall.

- Is there a relationship between the “swelling pressure” of a soil and the passive earth pressure?
- How much lateral movement might be anticipated as a result of the backfill becoming saturated?

1.2.6 Bearing Capacity for Shallow Foundations

The foundations for light structures are generally shallow spread footings (Fig. 1.13). The bearing capacity of the underlying (clayey) soils is computed based on the unconstrained compressive strength of the soil. Shallow footings can easily be constructed when the water table is below the elevation of the footings. In most cases, the water table is at a considerable depth, and the soil below the footing has a negative pore–water pressure. Undisturbed samples, held intact by negative pore–water pressures, are routinely tested in the laboratory. The assumption is made that the pore–water pressure conditions in the field will remain relatively constant with time, and therefore, the unconstrained compressive strength will also remain essentially unchanged. Based on this assumption, and a relatively high design factor of safety, the bearing capacity of the soil is computed.

The above design procedure has involved soils with negative pore–water pressures. It appears that the engineer has almost been oblivious to the problems related to the long-term retention of negative pore–water pressure when dealing with bearing capacity problems. Almost the opposite
INTRODUCTION TO UNSATURATED SOIL MECHANICS

attitude has been taken towards negative pore-water pressures when dealing with slope stability problems. That is, the attitude of the engineer has generally been that negative pore-water pressures cannot be relied upon to contribute to the shear strength of the soil on a long-term basis when dealing with slope stability problems. The two, seemingly opposite attitudes or perceptions, give rise to the question, "How constant are the negative pore-water pressures with respect to time?" Or, a more probing question might be, "Has the engineer's attitude towards negative pore-water pressures been strongly influenced by expediency?" This is a crucial question which requires further research and debate.

Other questions related to the design of shallow footings that might be asked are:

- What changes in pore-water pressures might occur as a result of sampling soils from above the water table?
- What effect does the _in situ_ negative pore-water pressure and a reduced degree of saturation have on the measured, unconfined compressive strength? How should the laboratory results be interpreted?
- Would _confined_ compression tests more accurately simulate the strength of an unsaturated soil for bearing capacity design?
- How much loss in strength could occur as a result of watering the lawn surrounding the building?

1.2.7 Ground Movements Involving Expansive Soils

There is no problem involving soils with negative pore-water pressures that has received more attention than the prediction of heave associated with the wetting of an expansive soil. Light structures such as a roadway or a small building are often subjected to severe distress subsequent to construction, as a result of changes in the surrounding environment (Figs. 1.14 and 1.15). Changes in the environment may occur as a result of the removal of trees, grass, and the excessive watering of a lawn around a new structure. The zone of soil undergoing volume change on an annual basis has been referred to as the "active zone." The higher the swelling properties of the soil, the greater will be the amount of heave to the structure.

It has been common practice to obtain undisturbed soil samples from the upper portion of the profile for one-dimensional oedometer testing in the laboratory. The laboratory results are used to provide quantitative estimates of potential heave. Numerous laboratory testing techniques and analytical procedures have been used in practice. Questions related to these procedures are discussed in detail in Chapter 13. However, other relevant questions that can be asked are:

- How much heave can be anticipated if the soil is flooded?
- How much heave can be anticipated if the negative pore-water pressures go to zero or remain at a slightly negative value?
- What is a reasonable final stress condition to assume for the pore-water pressures?
- What is the effect of prewetting or flooding the soil prior to constructing the foundation? How long must the prewetting continue? And what ground movements might continue subsequent to discontinuing the flooding?
- What is the effect of surcharging a swelling soil? How

Figure 1.13 Illustration of bearing capacity conditions for a light structure placed on soils with negative pore-water pressure.

Figure 1.14 Ground movements associated with the construction of shallow footings on an expansive soil (Hamilton, 1977).
1.3 TYPICAL PROFILES OF UNSATURATED SOILS

much might the potential heave be reduced by sur-
charging?
• What is the effect of placing an impervious membrane
around the perimeter of the footings?
• How much differential movement might one or more
large trees produce on the foundation?
• What are the satisfactory laboratory testing procedures
for measuring the swelling properties of an expansive
soil?

1.2.8 Collapsing Soils

In many respects, collapsing soils can be thought of as be-
having in an opposite manner to expansive soils. In both
the expansive and collapsing cases, the initial pore-water
pressures are negative. In both cases, movements are the
result of an increase in the negative pore-water pressure.
The wetting of a collapsing soil, however, results in a vol-
ume decrease. In this case, the soil is described as having
a metastable soil structure.

The collapse of the soil structure may occur within a man-
made or natural earth slope or in soil underlying a foun-
dation. Research associated with the behavior of collapsing
soils has been limited, and many questions remain to be
answered from both a research and practice standpoint:

• How should collapsing soils be tested in the labora-
tory?
• How should the laboratory data be interpreted and ap-
plied to practical problems?

1.2.9 Summary of Unsaturated Soils Examples

The above examples show that there are many practical
situations involving unsaturated soils that require an un-
derstanding of the seepage, volume change, and shear
strength characteristics. In fact, there is often an interaction
among, and a simultaneous interest in, all three of the as-
pcts of unsaturated soil mechanics. Typically, a flux
boundary condition produces an unsteady-state saturated/
unsaturated flow situation which results in a volume change
and a change in the shear strength of the soil. The change
in shear strength is generally translated into a change in
factor of safety. There may also be an interest in quanti-
fying the change of other volume-mass soil properties (i.e.,
water content and degree of saturation).

The classical one-dimensional theory of consolidation is
of central importance in saturated soil mechanics. The the-
ory of consolidation predicts the change in pore-water
pressure with respect to depth and time in response to a
change in total stress. The changes in pore-water pressure
are used to predict the volume change. The theory of con-
solidation does not play as important a role for unsatu-
rated soils as it does for saturated soils. The application of a
total stress to an unsaturated soil produces larger instantaneous
volume changes, but smaller volume changes with respect
to time. The induced pore-water pressures are consider-
ably smaller than the applied total stress. The more com-
mon boundary condition for unsaturated soils is a change
in flux as opposed to a change in total stress for a satu-
rated soil. Nevertheless, the theory of consolidation for unsat-
urated soils plays an important phenomenological role. It
assists the engineer in visualizing complex mechanisms,
providing a qualitative "feel" for the behavior of an unsat-
urated soil.

1.3 TYPICAL PROFILES OF UNSATURATED
SOILS

The microclimatic conditions in an area are the main factor
causing a soil deposit to be unsaturated. Therefore, unsat-
urated soils or soils with negative pore-water pressures can
occur in essentially any geological deposit. An unsaturated
soil could be a residual soil, a lacustrine deposit, a bedrock
formation, and so on. However, there are certain geologi-
10 1 INTRODUCTION TO UNSATURATED SOIL MECHANICS

tical categories of soils with negative pore–water pressures that have received considerable attention in the research literature. A few examples will illustrate some of the features common to these deposits.

1.3.1 Typical Tropical Residual Soil Profile

Tropical residual soils have some unique characteristics related to their composition and the environment under which they develop. Most distinctive is the microstructure which changes in a gradational manner with depth (Vargas, 1985; Brand, 1985). The in situ water content of residual soils is generally greater than its optimum water content for compaction. Their density, plasticity index, and compressibility are likely to be less than corresponding values for temperate zone soils with comparable liquid limits. Their strength and permeability are likely to be greater than those of temperate zone soils with comparable liquid limits (Mitchell and Sitar, 1982).

Most classical concepts related to soil properties and soil behavior have been developed for temperate zone soils, and there has been difficulty in accurately modeling procedures and conditions to which residual soils will be subjected. Engineers appear to be slowly recognizing that residual soils are generally soils with negative in situ pore–water pressures, and that much of the unusual behavior exhibited during laboratory testing is related to a matric suction change in the soil (Fredlund and Rahardjo, 1985).

A typical deep, tropical weathering profile is shown in Fig. 1.16 (Little, 1969). Boundaries between layers are generally not clearly defined. Numerous systems of classification have been proposed based primarily on the degree of weathering and engineering properties (Deere and Patton, 1971; Tuncer and Lohnes, 1977; Brand, 1982).

Zones of completely weathered or highly weathered rock that contain particulate soil but retain the original rock structure are termed saprolite. Once the deposit has essentially no resemblance of the parent rock, it is termed a lateritic or residual soil.

Figure 1.17 shows the profile and soil properties for a porous, saprolite soil from basalt in Brazil (Vargas, 1985).

![Figure 1.16 Schematic diagram showing a typical tropical residual soil profile (from Little, 1969).](image)

![Figure 1.17 Porous saprolite soil from basalt near Londrina, Brazil (from Vargas, 1985).](image)
1.3 TYPICAL PROFILES OF UNSATURATED SOILS

Matric suction, \( (u_s - u_w) \) (kPa)  

Water content, \( w \) (%)

<table>
<thead>
<tr>
<th>Depth (m)</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Brown, fine, dry, sand with roots</td>
<td>0.46</td>
<td>0.91</td>
<td>3.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yellow, silty sand</td>
<td>3.06</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Khaki shale with iron concretions</td>
<td>3.51</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Khaki, slickensided shale friable with iron concretions</td>
<td>3.81</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Khaki, slickensided shale friable with iron concretions and red sand</td>
<td>5.18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 1.18 Description of the soil profile at Welkom, South Africa, along with suction and water content profiles (from de Brujin, 1965).

The region has a hot, humid summer and a mild, dry winter climate, with an annual rainfall of less than 1500 mm. The structure is highly porous, and in some cases may be unstable, resulting in collapse upon saturation. The soil deposit is unsaturated, and the in situ pore-water pressure is negative.

1.3.2 Typical Expansive Soils Profile

Expansive soils deposits may range from lacustrine deposits to bedrock shale deposits. In general, expansive soils have a high plasticity (i.e., high liquid limit) and are relatively stiff or dense. The above description is typical, but not exclusive. The expansive nature of the soil is most obvious near ground surface where the profile is subjected to seasonal, environmental changes. The pore-water pressure is initially negative and the deposit is generally unsaturated. These soils often have some montmorillonite clay mineral present. The higher the amount of monovalent cations absorbed to the clay mineral (e.g., sodium), the more severe the expansive soils problem.

Expansive soils deposits and their related engineering problems have been reported in many countries. One of the first countries to embark on research into expansive soils was South Africa. Figure 1.18 shows a typical expansive soil profile, along with water content and soil suction values from near Welkom, South Africa (de Brujin, 1965). The area is well known for its foundation problems, giving rise to extensive damage to buildings and roads.

The soil profile (Fig. 1.18) shows that the soil is wetter and the suctions are lower in the zone below a covered roadway than below an open field.

Figure 1.19 shows a soil profile of an expansive soil in Tel Aviv, Israel (Katzir, 1974). The upper portion of the excavation consisted of a highly plastic, slickensided, fissured clay to a depth of 8-10 m. The water table was at 14 m. The upper portion had a liquid limit of 60%, a plastic limit of 25%, and a shrinkage limit of 11%. These properties are typical of swelling clay profiles.

Figure 1.19 Profile of expansive soil from Tel Aviv, Israel (from Katzir, 1974).
Extensive areas of western Canada are covered by pre-glacial, lacustrine clay sediments that are known for their expansive nature. Regina, Saskatchewan is located in a semi-arid area where the annual precipitation is approximately 350 mm. A typical soil profile is shown in Fig. 1.20 (Fredlund, 1973). The average liquid limit is 75% and the average plastic limit is 25%. The shrinkage limit is typically 15%. The lacustrine clay is classified as a calcium montmorillonite. Buildings founded on shallow footings often experience 50–150 mm of movement subsequent to construction.

The above soil profiles are typical of conditions which are found in many parts of the world. In each case, the natural water contents are low and the pore-water pressures are negative. Throughout each season, and from season to season, the soil expands and contracts in response to changes in the environment.

1.4 NEED FOR UNSATURATED SOIL MECHANICS

The success of the practice of soil mechanics can be traced largely to the ability of engineers to relate observed soil behavior to stress conditions. This ability has led to the transferability of the science and a relatively consistent engineering practice. Although this has been true for saturated soils, it has not been the case for unsaturated soils. Difficulty has been experienced in extending classical soil mechanics to embrace unsaturated soils. This can be borne out by the empirical nature of much of the research associated with unsaturated soils.

The question can be asked: "Why hasn't a practical science developed and flourished for unsaturated soils?" A cursory examination may suggest that there is no need for such a science. However, this is certainly not the case when the problems associated with expansive soils are considered. Jones and Holtz (1973) reported that in the United States alone: "Each year, shrinking and swelling soils inflict at least $2.3$ billion in damages to houses, buildings, roads, and pipelines—more than twice the damage from floods, hurricanes, tornadoes, and earthquakes!" They also reported that 60% of the new houses built in the United States will experience minor damage during their useful lives and 10% will experience significant damage—some beyond repair.

In 1980, Krohn and Slosson estimated that $7$ billion are spent each year in the United States as a result of damage to all types of structures built on expansive soils. Snethen (1986) stated that, "While few people have ever heard of expansive soils and even fewer realize the magnitude of the damage they cause, more than one fifth of American families live on such soils and no state is immune from the problem they cause. Expansive soils have been called the "hidden disaster": while they do not cause loss of life, economically, they are one of the United States costliest natural hazards." The problem extends to many other countries of the world. In Canada, Hamilton (1977) stated that: "Volume changing clay subsoils constitute the most costly natural hazard to buildings on shallow foundations.

![Figure 1.20 Profile of expansive soils from Regina, Canada (from Fredlund, 1973).](image-url)
in Canada and the United States. In the Prairie Provinces alone, a million or more Canadians live in communities built on subsoils of very high potential expansion."

It would appear that, internationally, most countries in the world have problems with expansive soils. Many countries have reported their problems at research conferences. Some countries reporting problems with expansive soils are: Australia, Argentina, Burma, China, Cuba, Ethiopia, Ghana, Great Britain, India, Iran, Israel, Kenya, Mexico, Morocco, South Africa, Spain, Turkey, and Venezuela. In general, the more arid the climate, the more severe is the problem. A series of conferences have been held to help cope with the expansive soil problem. These are:

10. Sixth International Conference on Expansive Soils, New Delhi, India (1987).

There is also the need for reliable engineering design associated with compacted soils, collapsing soils, and residual soils. Soils that collapse usually have an initially low density. The soils may or may not be subjected to additional load, but they are usually given access to water. The water causes an increase in the pore-water pressures, with the resulting that the soil volume decreases. The process is similar to that occurring in an expansive soil, but the direction of volume change is the opposite. Examples of soil collapse have recently been reported in numerous countries. In the United States, for example, Johnpeer (1986) reported examples of collapse in New Mexico. Leaky septic tanks were a common initiating factor in the soil collapse.

Inhabited areas with steep slopes consisting of residual soils are sometimes the site of catastrophic landslides which claim many lives. A widely publicized case is the landslide at Po-Shan in Hong Kong which claimed 67 lives (The Commission of Inquiry, 1972). Similar problems have been reported in South American countries and other parts of the world.

The soils involved are often residual in genesis and have deep water tables. The surface soils have negative pore-water pressures which play a significant role in the stability of the slope. However, heavy, continuous rainfalls can result in increased pore-water pressures to a significant depth, resulting in the instability of the slope. The pore-water pressures along the slip surface at the time of failure may be negative or positive.

There appear to be two main reasons why a practical science has not developed for unsaturated soils (Fredlund, 1979). First, there has been the lack of an appropriate science with a theoretical base. This commences with a lack of appreciation of the engineering problems and an inability to place the solution within a theoretical context. The stress conditions and mechanisms involved, as well as the soil properties that must be measured, do not appear to have been fully understood. The boundary conditions for an analysis are generally related to the environment and are difficult to predict. Research work has largely remained empirical in nature, with little coherence and synthesis. There has been poor communication among engineers, and design procedures are not widely accepted and adhered to.

Second, there appears to be the lack of a system for financial recovery for services rendered by the engineer. In the case of expansive soil problems, the possible liability to the engineer is often too great relative to the financial remuneration. Other areas of practice are more profitable to consultants. The owner often reasons that the cost outweighs the risk. The hazard to life and injury is largely absent, and for this reason little attention has been given to the problem by government agencies. Although the problem basically remains with the owner, it is the engineer who has the greatest potential for circumventing possible problems.

Certainly, there is a need for an appropriate technology for unsaturated soil behavior. Such a technology must: 1) be practical, 2) not be too costly to employ, 3) have a sound theoretical basis, and 4) run parallel in concept to conventional saturated soil mechanics.

1.5 SCOPE OF THE BOOK

This book addresses the field of unsaturated soil mechanics. An attempt is made to cover all of the aspects normally associated with soil mechanics. When the term "unsaturated soil mechanics" is used, the authors are referring to soils which have negative pore-water pressures.

The aspects of interest to geotechnical engineering fall into three main categories. These can be listed as problems related to: 1) the flow of water through porous media, 2) the shear strength, and 3) the volume change behavior of unsaturated soils. No attempt is made to duplicate or redevelop information already available in classic saturated
soil mechanics books. This book should be used to assist the geotechnical engineer in understanding soil mechanics concepts unique to unsaturated soils. At the same time, these concepts have been developed and organized to appear as logical and simple extensions of classical saturated soil mechanics. Subjects such as clay mineralogy and physico-chemical properties of soils are vitally important to understanding why soils behave in a certain manner. However, the readers are referred to excellent references for coverage of these subjects (Yong and Warkentin, 1966; Mitchell, 1976).

Most soil mechanics problems can be related to one (or more) of the three main soil properties. These categories are 1) the coefficient of permeability, 2) the shear strength parameters, and 3) the volume change indices. There are three chapters written to cover each category of soil properties. The chapters can be described as: 1) the theory related to the soil property, 2) the measurement of the soil property, and 3) the application of the soil property to one or more soil mechanics problems. Chapters 5, 6, and 7 present the theory, measurement, and application, respectively, relative to the air and water flow. Chapters 9, 10, and 11 present the theory, measurement, and application, respectively, relative to shear strength. Chapters 12, 13, and 14 present the theory, measurement, and application, respectively, relative to volume change. The equipment required and details on the testing procedures are presented under the "measurement" chapters in each case.

Considerable attention is devoted to the description of the stress state variables required to describe unsaturated soil behavior (Chapter 3). The space devoted to this topic is in keeping with the importance of this fundamental concept. Chapter 4 describes the techniques and devices available to measure soil suction. This is an area of ongoing research, and an attempt is made to present the most recent information on the measurement of total, matric, and osmotic suctions.

The concept, theory, and application of pore pressure parameters is presented in Chapter 8. Pore pressure parameters involve the combination of the constitutive relations for the soil and the phase relations (Chapter 2) to predict the change in pore-air and pore-water pressures as a result of applying a total stress and not allowing drainage.

The theory of consolidation, as well as unsteady-state flow analyses, combine the volume change properties and the flow laws of a soil. The one-dimensional consolidation theory is presented in Chapter 15, while two- and three-dimensional, unsteady-state flow is presented in Chapter 16.

The main objective of this book is to synthesize theories associated with the behavior of unsaturated soils. The theoretical derivations are presented in considerable detail because unsaturated soil behavior is a relatively new area of study and many of the derivations are not readily available to engineers. There is a need for case histories, and it is anticipated that these will become more common in future decades. Hopefully, as the analyses are related to case histories, the engineer can benefit from the consistent theoretical context provided by this book.

1.6 PHASES OF AN UNSATURATED SOIL

An unsaturated soil is a mixture of several phases. It is important to establish the number of phases comprising the soil since it has an influence on how the stress state of the mixture is defined. First, it is important to define what is meant by a phase. On the basis of the definition of a phase, it is proposed that an unsaturated soil actually consists of four phases rather than the commonly referred to three phases. It is postulated that in addition to the solid, air, and water phases, there is the air-water interface that can be referred to as the contractile skin. Let us pursue the justification for reference to the contractile skin as an independent phase.

1.6.1 Definition of a Phase

In order for a portion of a mixture to qualify as an independent phase, it must have: 1) differing properties from the contiguous materials, and 2) definite bounding surfaces. These two conditions must be met in order to identify an independent phase. It is easy to understand how a saturated soil consists of two phases (i.e., soil solids and water). It is also quite understandable that the air becomes another independent phase for an unsaturated soil. Each of these phases (i.e., soil solids, water, and air) obviously meet the requirements for designation as a phase.

It is also possible for a phase to change state, as is the case when water freezes. The ice becomes an independent phase from the water. However, more important is the examination of the properties and extent of the air-water interface.

1.6.2 Air-Water Interface or Contractile Skin

The most distinctive property of the contractile skin is its ability to exert a tensile pull. It behaves like an elastic membrane under tension interwoven throughout the soil structure. It appears that most properties of the contractile skin are different from those of the contiguous water phase (Davies and Rideal, 1963). For example, its density is reduced, its heat conductance is increased, and its birefringence data are similar to those of ice. The transition from the liquid water to the contractile skin has been shown to be distinct or jumpwise (Derjaguin, 1965). It is interesting to note that insects such as the "water spider" walk on top of the contractile skin, and those such as the "backswimmer" walk on the bottom of the contractile skin.
Numerous approaches could be taken in the development of the discipline of unsaturated soil mechanics. The approaches may range from an empirical approach based strictly on experience to a particulate mechanics or quantum mechanics approach. The approach used throughout this book can be referred to as a macroscopic, phenomenological approach to unsaturated soil behavior. In other words, the science is developed around observable phenomena, while adhering to continuum mechanics principles. This approach has proven to be most successful in saturated soil mechanics and should be retained in unsaturated soil mechanics. An attempt is made to ensure a smooth transition in rationale between the saturated and unsaturated cases.

The authors are reticent to introduce new variables and terminology to unsaturated soil mechanics. However, the terminology associated with saturated soil mechanics is not sufficient when applied to unsaturated soil mechanics. As a result, a few more universally acceptable terms are proposed. These terms are common to continuum mechanics and are defined within a thermodynamic context. The following definitions, put in a succinct form, are based on numerous continuum mechanics and thermodynamic references:

1) **State**: Nonmaterial variables required for the characterization of a system.

2) **Stress state variable**: The nonmaterial variables required for the characterization of the stress condition.

3) **Deformation state variables**: The nonmaterial variables required for the characterization of deformation conditions or deviations from an initial state.

4) **Constitutive relations**: Single-valued equations expressing the relationship between state variables.

*The International Dictionary of Physics and Electronics* (Michels, 1961) defines state variables as:

"a limited set of dynamical variables of the system, such as pressure, temperature, volume, etc., which are sufficient to describe or specify the state of the system completely for the considerations at hand."

Fung (1965) describes the state of a system as that "information required for a complete characterization of the system for the purpose at hand." Typical state variables for an elastic body are given as those variables describing the strain field, the stress field, and its geometry. The state variables must be independent of the physical properties of the material.

Constitutive relations, on the other hand, are single-valued expressions which relate one state variable to one or more other state variables (Fung, 1969). A stress versus strain relationship is a constitutive relation which describes the mechanical behavior of a material. The material properties involved may be an elastic modulus and a Poisson's ratio. The ideal gas equation relates pressure to density, and temperature and is called a constitutive equation. The gas constant is the material property. Simple, idealized constitutive equations are well established for a nonviscous
1.8 HISTORICAL DEVELOPMENTS

The first ISSMFE conference (International Society for Soil Mechanics and Foundation Engineering) in 1936 provided a forum for the establishment of principles and equations relevant to saturated soil mechanics. These principles and equations have remained pivotal throughout subsequent decades of research. This same conference was also a forum for numerous research papers on unsaturated soil behavior. Unfortunately, a parallel set of principles and equations did not immediately emerge for unsaturated soils. In subsequent years, a science and technology for unsaturated soils has been slow to develop (Fredlund, 1979). Not until the research at Imperial College in the late 1950's did the concepts for understanding unsaturated soils behavior begin to be established (Bishop, 1959). The research of Lytton (1967) in the United States did much to ensure that the understanding of unsaturated soil behavior was founded upon a sound theoretical basis. Namely, that all theories were consistent with the principles put forth in continuum mechanics. The following is a brief history, highlighting the early developments in our understanding of the behavior of unsaturated soils. Most of the early research on unsaturated soils was related to the flow of water in the zone of negative pore-water pressure (i.e., capillary flow).

Associated with the urban development in the 1930's was the construction of numerous engineering works such as irrigation and transportation projects. One of the first problems that appeared to perplex civil engineers was that of the movement of water above the groundwater table. The term "capillarity" was adopted to describe the phenomenon of water flow upward from the static groundwater table. This term was selected because of the similarity to the operation of a capillary tube. Example problems were illustrated by Hogentogler and Barber in 1941. In their first example [Fig. 1.22(a)], water was shown to move up and over the impervious core of a dam, even when the core extended above the reservoir elevation. Water was reported to move over the impervious wall and result in seepage problems along the downstream face of the dam.

A second example, Fig. 1.22(b), illustrated the ineffectiveness of cutoff ditches in intercepting groundwater flow. The freezing and thawing of capillarity water lead to embankment instability and subgrade failures. These problems initiated research on soil capillarity for at least two decades.

Two research papers on soil capillarity were presented by Ostashev at the ISSMFE conference in 1936. Numerous factors, including pore-water pressure and capillary force, were described as affecting the mechanism of capillary flow. The Fourier heat flow equation was proposed for modeling the moisture flow process. At the same conference, two apparatuses were proposed for measuring the capillary pressure and capillary rise of water in soils (Boulanchev, 1936).

Hogentogler and Barber (1941) attempted to give a comprehensive review of the nature of the capillary water. It was suggested that capillary water responded in accordance with the capillary rise equation for a fine bore tube. A capillarimeter was built to study the capillary phenomenon. The apparatus was a modification of the Bartell cell originally proposed in 1896 for use in studying lime and cement mortars. The capillarimeter measured the air entry value of the soil. It became the primary apparatus used for studying soil capillarity for several decades.

The phenomenon of capillary flow was illustrated using a simple beaker and sand-filled tube as shown in Fig. 1.23. Hogentogler and Barber (1941) offered the following comment on capillary flow, stating that, "... it is considered that capillary moisture of this type conforms strictly to recognized concepts of surface tension, the force of gravity and the principles of hydraulics as applied to free water..." It was also suggested that the strength of an
unsaturated soil was greatly influenced by the state of stress in the capillary water. The state of stress in the capillary water was also related to evapotranspiration and the relative humidity of the air above the soil. It was suggested that the "stabilizing effect of capillary saturation" could be used in a practical way to stabilize the face of an excavation. Possible, practical concepts and applications associated with negative pore-water pressures appear to have been well developed by these researchers.

Terzaghi (1943) in his book, *Theoretical Soil Mechanics*, summarized the work of Hogentogler and Barber (1941) and endorsed the concepts related to the capillary tube model. The importance of the air-water interface was emphasized with respect to its effect on soil behavior. An equation was derived for the time required for the rise of water in the capillary zone. The assumption was made that the porosity, $n$, and the coefficient of permeability, $k$, were constants.

$$ t = \frac{nh_c}{k} \left[ \log \left( \frac{h_c}{h_c - z} \right) - \frac{z}{h_c} \right] \quad (1.1) $$

where:

- $t$ = time
- $n$ = porosity
- $h_c$ = maximum capillary rise
- $z$ = vertical distance above the groundwater table, corresponding to the elapsed time, $t$.

Valle-Rodas (1944) performed open tube and capillarimeter tests on uniform sands. In the open tests, sand was placed in a glass tube which had one end immersed in water. Experimental results showed an uneven water content in the sand in the capillary zone (Fig. 1.24). Moreover, it was found that there was a hysteresis in water content with respect to wetting and drying.

Further open tube and capillarimeter tests were performed by Lane and Washburn (1946) on cohesionless soils ranging from silts to gravels. The results indicated reasonable agreement between the measured capillary rise and the values predicted by Terzaghi's equation for the height of capillary rise. Terzaghi's equation for the prediction of the rate of capillary rise [Eq. (1.1)] appeared to overestimate the rate of capillary rise (Fig. 1.25). It was postulated that the discrepancy was due primarily to changing permeability in the capillary zone. It was shown that reduced values for the coefficient of permeability gave better correlations. Krynine (1948) reanalyzed the results of many capillary tests run between 1934 and 1946 and came to a similar conclusion. His results are summarized in Fig. 1.26.

Sitz (1948) noted that capillary water would rise to more than 10 m above a static groundwater table. He suggested that capillary water be subdivided into gravitational and molecular water. It was postulated that gravitational capillary water had properties similar to ordinary water, while molecular capillary water was assumed to have unique
properties. The property of water which was of primary interest was its ability to withstand high tensile stresses without cavitating or boiling.

Bematzik (1948) observed an increase in the strength of a soil as a result of the air–water menisci. He suggested the use of an unconfined compression type of test on a soil specimen in order to study capillary tension.

Lambe (1951a) performed open tube capillary rise and drainage tests on graded sands and silts with various initial degrees of saturation. Negative pore–water pressures were measured at various locations along the specimens (Figs. 1.27 and 1.28). He again noted that the degree of saturation in the capillary zone was not 100%, and that the soil property controlling flow was noted to be different from the saturated coefficient of permeability. It was concluded that the hydraulic gradient was not uniform across the capillary zone. A large part of the available head appeared to be lost in the zone immediately behind the wetting front.

The historical review of the period up to the 1950’s shows that most of the attention given to unsaturated soils was related to capillary flow. An attempt was made to use the capillary tube rise model (see Chapter 4) to explain the observed phenomenon. Although this model was of some value, it had limitations which became increasingly obvious. In fact, attempts to totally rely on the capillary tube rise model appear to be a significant factor in the slow development of unsaturated soil mechanics.

Research into the volume change and shear strength of unsaturated soils commenced with new impetus in the late 1950’s. Some of the researchers were Black and Cmney (1957), Bishop et al., (1960), Aitchison (1967), and Williams (1957). The research resulted in the proposal of several so-called effective stress equations for unsaturated soils. During the next decade, reservations were expressed regarding the use of a single-valued effective stress equation. There has subsequently been a slow change towards the acceptance of two independent stress state variables (Coleman, 1962; Matyas and Radhakrishna, 1968; Fredlund and Morgenstern, 1977).

One of the pioneers to strongly advocate a sound theo-
retical basis for unsaturated soils' theories was Lytton (1967). His research drew upon the mixtures theories in continuum mechanics. These principles were applied to the multiphase, unsaturated soil system. Excellent direction was given for future research.

There are separate sections in this book devoted to a review of the historical developments related to each of the fundamental properties common to unsaturated soil mechanics. Details concerning the historical development towards the use of independent stress state variables are given in Chapter 3. The development and application of saturated-unsaturated flow modeling, subsequent to emphasis on the capillary model, are presented in Chapter 5. Developments in the area of the shear strength of unsaturated soils are presented in Chapter 9. Similar developments in the area of volume change in unsaturated soils are given in Chapter 12. In each of the above areas, much of the original work resulted in a somewhat semi-empirical approach towards understanding the behavior of unsaturated soils. With time, the principles of continuum mechanics that were found to be successful in saturated soil mechanics have also become the basis for unsaturated soil mechanics. It is on this infant, but sound, theoretical basis that the theory in this book has been assembled.

Figure 1.28 Degree of saturation versus distance above the bottom of a soil column (from Lambe, 1951a).