1. Significance of $f''(a)$ when $f'(a)$ vanishes:

   Taylor's theorem for the function f at the point a, of second order, announce that

   $$f(a+h) = f(a) + f'(a)h + f''(a)h^2 + R$$

   for a negligible remainder R. This means that if $f'(a) = 0$ then $f(a+h) = f(a) + f''(a)h^2 + R$ or that $f(a+h) - f(a) = f''(a)h^2 + R$. This is the second derivative test, which claims: if $f''(a) < 0$ then $f(a+h) - f(a) < 0$ or that $f(a+h) < f(a)$ for all h. This means that a is a local max for f. Similarly, if $f''(a) > 0$ then $f(a+h) - f(a) > 0$ for all h, which means that $f(a+h) > f(a)$ for all h, that is, f has a local min at the point a

2. However if the second derivative vanishes at a then the analysis of the local max/min is determined by the next term of the Taylor polynomial (the next term is of degree 3)

   $$f(a+h) = f(a) + f'(a)h + f''(a)h^2 + f^{(3)}(a)h^3 + R$$

   so in the absence of $f'(a)$ and $f''(a)$ we have $f(a+h) - f(a) = f^{(3)}(a)h^3 + R$. But now since the power of h is 3 then there for $h > 0$ the expression $f(a+h) - f(a)$ has one sign, while for $h < 0$ the expression $f(a+h) - f(a)$ has the opposite sign. This means that the function cannot have a local max nor a local min at the point a

   An example of such situation is $f(x) = x^3$ at $a = 0$. $f''(x) = 6x$, and $f''(0) = 0$. Now $f^{(3)}(x) = 6$, so $f(0+h) - f(0) = f^{(3)}(0)h^3 = 6h^3$. It is clear that $f(0+h) - f(0)$ is positive when $h > 0$ and it is negative when $h < 0$.

3. Note that if $f^{(3)}(a) = 0$ as well, then the decision about local max/min is postponed even further, to be decided by the value of $f^{(4)}(a)$ because

   $$f(a+h) = f(a) + f'(a)h + f''(a)h^2 + f^{(3)}(a)h^3 + f^{(4)}(a)h^4 + R,$$

   which, in the absence of the first three derivatives becomes

   $$f(a+h) - f(a) = f^{(4)}(a)h^4 + R,$$

   and now it is possible to make a decision again because the power of h is 4. Example $f(x) = x^4$.

4. This idea is extended to the function of several variable through the Hessian matrix, that is the "second derivative matrix". First note that for cases of $n > 1$ all the elements are vectors and matrices: the values of $f(\boldsymbol{a})$ and $f(\boldsymbol{a})$ are numbers, $\boldsymbol{h}$ and ∇f are $n \times 1$ vectors, whereas the second derivatives $H(\boldsymbol{a})$ $n \times n$ matrices. (Question: What does a third derivative look like? Do you know of any mathematical objects that could represent these derivatives? Have we come to our mathematical limits with the Taylor polynomials of degree 2!!?)

   so the Taylor's polynomial of degree 2 will look like (this is a version of (2.80) in page 96 consistent with linear algebra notations we have had:

   $$f(\boldsymbol{a} + \boldsymbol{k}) = f(\boldsymbol{a}) + \nabla f(\boldsymbol{a}) \cdot \boldsymbol{k} + \boldsymbol{k}^T H(\boldsymbol{a})\boldsymbol{k} + R_{\boldsymbol{a},2}.$$

So when $\nabla f(\boldsymbol{a}) = 0$ then

$$f(\boldsymbol{a}+\boldsymbol{k}) = f(\boldsymbol{a}) + \boldsymbol{k}^T H(\boldsymbol{a})\boldsymbol{k} + R_{\boldsymbol{a},2}.$$

Thus, the component that really approximate the value of $f(\boldsymbol{a}+\boldsymbol{k})$ is $\boldsymbol{k}^T H(\boldsymbol{a})\boldsymbol{k}$. Now, in analogy with item 1 above, to decide whether the point $\boldsymbol{a}$ is a local max or a local min, we need to know if the difference $f(\boldsymbol{a}+\boldsymbol{k}) - f(\boldsymbol{a})$ is positive or negative; hence we should study the sign of $\boldsymbol{k}^T H(\boldsymbol{a})\boldsymbol{k}$.

5. A $n \times n$ matrix A is said to be 'positive definite' if for any n vector $\boldsymbol{v}$ the expression $\boldsymbol{v}^T A\boldsymbol{v}$ is always positive (this is analogical to the $hf''(a)h = f''(a)h^2 > 0$.) Similarly, A is said to be 'negative definite' if for any n vector $\boldsymbol{v}$ we have $\boldsymbol{v}^T A\boldsymbol{v} < 0$.

   Check that the following matrices are positive definite:

$$A = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 3 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 2 \\ 2 & 5 \end{bmatrix} \quad C = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 5 & 1 \\ 3 & 1 & 36 \end{bmatrix}$$

   While matrix A is trivially positive definite, to verify matrix B is positive definite you may need to write the expression $a^2 + 4ab + 5b^2$ as $(a+2b)^2 + b^2$. However it is a very difficult task to verify that matrix C is positive definite. In the next few items we shall see some alternative methods for deciding when a matrix is positive definite.

6. Verify that the following matrices are NOT positive definite nor negative definite (find two vectors $\boldsymbol{v}$ and $\boldsymbol{w}$ such that the result of $\boldsymbol{v}^T A\boldsymbol{v}$ is positive and the result of $\boldsymbol{w}^T A\boldsymbol{w}$ is negative):

$$A = \begin{bmatrix} 2 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 3 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix} \quad C = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 5 & 1 \\ 3 & 1 & 34 \end{bmatrix}$$

   Again the matrix C presents us with a challenge.

7. Geometric interpretation of a 'symmetric positive definite matrix': any matrix M defines a linear transformation operation, and in particular and $n \times n$ matrix defines a linear operator from $\mathbb{R}^n \longrightarrow \mathbb{R}^n$ that causes some distortion in $\mathbb{R}^n$. Next, let $\boldsymbol{v}$ be vector in $\mathbb{R}^n$ and note that under transformation of M we have a new vector $M\boldsymbol{v}$ whose magnitude is now changes to

$$||M\boldsymbol{v}|| = \sqrt{(M\boldsymbol{v})^T(M\boldsymbol{v})} = \sqrt{\boldsymbol{v}^T M^T M\boldsymbol{v}} = \sqrt{\boldsymbol{v}^T A\boldsymbol{v}}$$

   where $A = M^T M$ is a symmetric matrix. Of course if A is positive definite then this new norm will be positive for each vector $\boldsymbol{v}$, else this linear transformation (defined by M) makes no sense as a 'real' transformation because it may make the vectors have negative norm (which is a very questionable transformation.) Can you determine the matrix M corresponding to the matrix A as in item 5 above.

8. Historically the Special theory of relativity uses a transformation in the form

$$M = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & ic \end{bmatrix}$$

where i is the imaginary number with $i^2 = -1$ and c denotes the speed of light. Note that this matrix in action will transform the space-time vectors (x, y, z, t) (three dimensions of space and one of time) in such manner that the norm of a photon in the 4 dimensional space-time will become

$$[x\ y\ z\ t]\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -c^2 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \\ t \end{bmatrix} = x^2 + y^2 + z^2 - c^2t^2 = 0$$

therefore a photon instantly travels from point to point! Since this matrix is not positive definite then it divides the vectors in the space-time into three groups: those with norm greater than 0 (known as spacelike) and those with norm less than 0 (known as timelike) and finally those with norm equal to zero (which is known as lightlike or belonging to those objects who travel with the speed of light.)

9. Although our definition of 'positive definite' is very intuitive and useful in terms of analysis of the nature of critical points, as you have seen in items 5 and 6 it is next to impossible to use our definition of 'positive definite' to decide which matrices are positive definite. Our textbook (see the last equation on page 96) uses the definition of eigen values and eigen vectors of the matrix $H(\boldsymbol{a})$. If the vectors of the increment ($\boldsymbol{k}$) are along the the eigen vectors of $H(\boldsymbol{a})$ then $H(\boldsymbol{a})\boldsymbol{k}$ reduces to $\lambda\boldsymbol{k}$. In this case the expression $\boldsymbol{k}^T H(\boldsymbol{a})\boldsymbol{k}$ becomes $\lambda\boldsymbol{k}^T\boldsymbol{k}$. This argument presents a relation between positive definiteness and the sign of the eigen values of the matrix $H(\boldsymbol{a})$. Read more about this idea in page 96 and 97. Here we use an alternative characterization of positive definiteness, known as 'principle minor' method:

   a) A **Principle minor** of a real symmetric matrix is the minor of a the determinant corresponding to a diagonal element of the matrix.

   b) 'An increasing sequence of principle minors' is a sequence of minors that start with a 1 minor, then a 2×2 minor, and then a 3×3 minor, etc. that is

$$det(A) = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$$

   has the following list of increasing sequences of principle minors:

$$a_{11}, \quad \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}, \ det(A)$$

   and here is another such sequence:

$$a_{22}, \quad \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix}, \ det(A)$$

   As an exercise try to write all the principle minors of the matrix C of item 5, and of item 6.

10. The following equivalences forms some characterization of positive definiteness for real symmetric matrices. The following are equivalent: (accept without proof)

   a) A real symmetric matrix A is positive definite

b) there is an increasing sequence of minors all of which are positive

c) every increasing sequence of minors have all terms positive

Similarly thee is the following characterization of negative definiteness (note that if A is positive definite then $-A$ is negative definite.) The following are equivalent: (proof left as an exercise)

a) A is negative definite

b) one particular increasing sequence of minors has odd terms negative and even terms positive,

c) all the increasing sequences of principle minors have odd terms negative and even terms positive.

11. Conclusion:

    a) A matrix is NOT positive definite iff there exists one principle minors which is negative.

    b) And similarly, a matrix A is NOT negative definite iff there is some principle minor of odd size that is positive (OR there is some principle minor of even size that is negative.)

    c) This latest characterization leads to the following result: A matrix A is neither positive definite nor negative definite iff there exists a principle minor of even size less than zero (negative).

    d) A Saddle point is a point where two events of local max and local min take place concurrently. This means that there are directions along which "the second derivative is negative" and there are directions along which "the second derivative is positive". In our language of minors this translates to: there is one principle minor (of odd size) that is negative, AND either one of the following: (1) there is a principle minor of odd size that is positive, OR there is a principle minor of even size that is negative.

12. Read theorem 2.82 and compare with the 0previous item.