MAT 237: Supplementary notes on optimization

In these notes we provide a little more information about optimization than can be found in the basic online notes for MAT237.

Recall that if $f : \mathbb{R} \to \mathbb{R}$ is differentiable, then

$$\text{if } a \text{ is a local max or local min for } f, \text{ then } f'(a) = 0.$$

If f is C^2, then we can say more:

$$\left.\begin{array}{l} f'(a) = 0 \\ f''(a) > 0 \end{array}\right\} \implies a \text{ is a local min } \implies \left\{\begin{array}{l} f'(a) = 0 \\ f''(a) \geq 0 \end{array}\right.$$

Similarly,

$$\left.\begin{array}{l} f'(a) = 0 \\ f''(a) < 0 \end{array}\right\} \implies a \text{ is a local max } \implies \left\{\begin{array}{l} f'(a) = 0 \\ f''(a) \leq 0 \end{array}\right.$$

These facts remain true if the domain of f is any open subset of $\mathbb{R}$.

All of these facts have parallels for multi-variable calculus. We summarize these in the following theorem.

Theorem 1. *Assume that U is an subset of $\mathbb{R}^n$ and that $\mathbf{a} \in U^{int}$.*

1. *If $f : U \to \mathbb{R}$ is differentiable, then*

$$\textit{if } \mathbf{a} \textit{ is a local max or local min for } f, \textit{ then } \nabla f(\mathbf{a}) = 0.$$

2. *If f is C^2, and $\mathbf{a}$ is a local min for f, then in addition,*

$$\textit{all eigenvalues of } H(\mathbf{a}) \ \geq 0,$$

and if $\mathbf{a}$ is a local max for f, then

$$\textit{all eigenvalues of } H(\mathbf{a}) \ \leq 0.$$

Here $H(\mathbf{a})$ denotes the Hessian matrix of f, evaluated at the point $\mathbf{a}$.

3. *If f is C^2, then*

$$\textit{if} \quad \left\{\begin{array}{c} \nabla f(\mathbf{a}) = 0 \textit{ and} \\ \textit{all eigenvalues of } H(\mathbf{a}) < 0, \end{array}\right. \quad \textit{then } \mathbf{a} \textit{ is a local max for } f$$

$$\textit{if} \quad \left\{\begin{array}{c} \nabla f(\mathbf{a}) = 0 \textit{ and} \\ \textit{all eigenvalues of } H(\mathbf{a}) > 0, \end{array}\right. \quad \textit{then } \mathbf{a} \textit{ is a local min for } f$$

To avoid unnecessary repetition, and because basic ideas in the "max" and "min" cases are identical, we will focus on minimizers in what follows.

Example 1. It can happpen that all eigenvalues of $H(\mathbf{a})$ are nonnegative, but $\mathbf{a}$ is not a local min. An example is $f : \mathbb{R}^n \to \mathbb{R}$ defined by $f(\mathbf{x}) = -\|\mathbf{x}\|^3$. For this example, it is a straightforward exercise to check that $H(\mathbf{0}) = 0$ (that is, the $n \times n$ matrix whose entries are all 0), but f does not have a local min at $\mathbf{0}$, since $0 = f(\mathbf{0}) > f(\mathbf{x})$ for every $\mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$.

It can also happen that none of the eigenvalues of $H(\mathbf{a})$ are strictly positive, and yet f still has a local min at $\mathbf{a}$. The example is essentially the same — consider $f(\mathbf{x}) = \|\mathbf{x}\|^3$. This function again satisfies $H(\mathbf{a}) = 0$, but now it has a local min at $\mathbf{0}$.

Example 2. For any distinct positive real numbers, say $0 < a < b$, consider the elliptic paraboloid defined by the equation $z = ax^2 + by^2$ (pictured in Figure 1 below for $a = 5$, $b = 6$). At any height h, let $\rho_h(x, y)$ denote the square of the distance from the point $(0, 0, h)$ on the z-axis to the point $(x, y, f(x, y))$ on the graph of the paraboloid. That is, $\rho_h \colon \mathbb{R}^2 \to \mathbb{R}$ is given by $\rho_h(x, y) = x^2 + y^2 + (f(x, y) - h)^2$. As we will show below, for any h, the function ρ_h has a critical point at $(0, 0)$. In this example we will determine the nature of this critical point as h varies.

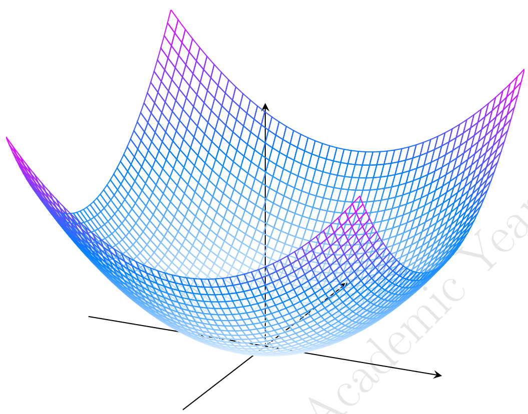
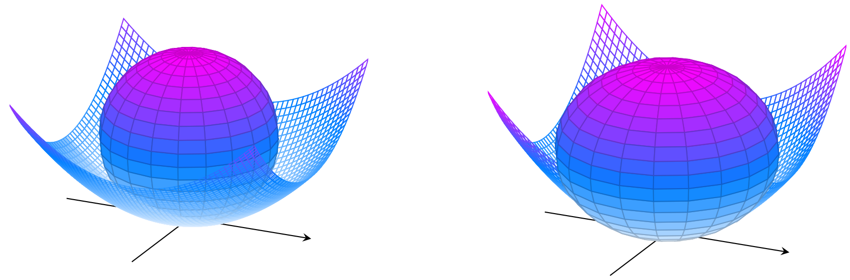


FIGURE 1. The paraboloid $z = 5x^2 + 6y^2$

Before solving the problem analytically, let us attempt to gain a geometric understanding of what we should expect for the result by considering the level sets of the functions ρ_h. We can picture the level sets of ρ_h by graphing spheres of radius $\sqrt{R}$ centred at the point $(0, 0, h)$ on the same set of axes as we graph the paraboloid $z = ax^2 + by^2$: the intersection is exactly the level set $\rho_h = R$. Thus the level set containing the critical point $(0, 0)$ is obtained by intersecting the sphere of radius h centred at $(0, 0, h)$ with the paraboloid. By taking h small enough so that the cross-sections of the sphere (circles) lie in the interior of the cross-sections of the paraboloid (ellipses), we can ensure that any point in a neighbourhood of $(0, 0)$ lies outside the sphere, causing the point $(0, 0)$ to be a local *minimum* for ρ_h (see Figure 2a). Conversely, if we take the h large enough so that the cross-sections of the paraboloid lie inside the cross-sections of the sphere, then $(0, 0)$ will be a local *maximum* for ρ_h (see Figure 2b). In between these two extremes, we might expect that the cross-sections of the sphere fit inside the cross sections of the paraboloid in the x-direction, but not in the y-direction. Thus if one moves slightly away from the origin in the x-direction, one ends up in a higher level set, while if one moves slightly away in the y-direction one ends up in a lower level set. For such values of h, the origin looks like a saddle point!

Now let us is see if we can support our geometric intuition using Theorem 1. Using the chain rule, we find that

$$\partial_x \rho_h(x, y) = 2(x + \partial_x f(x, y)[f(x, y) - a]) = 2x(1 + 2a[ax^2 + by^2 - h]),$$
$$\partial_y \rho_h(x, y) = 2(y + \partial_y f(x, y)[f(x, y) - a]) = 2y(1 + 2b[ax^2 + by^2 - h]),$$



(A) Small h.

(B) Large h.

FIGURE 2. Nature of the critical point as h varies.

which confirms that $(x, y) = (0, 0)$ is indeed a critical point of ρ_h for any height h. Differentiating once more, we find

$$\begin{aligned}
\partial_{xx}\rho_h(x, y) &= 2 + 4a(ax^2 + by^2 - h) + 8a^2x^2,\\
\partial_{yy}\rho_h(x, y) &= 2 + 4b(ax^2 + by^2 - h) + 8b^2y^2,\\
\partial_{xy}\rho_h(x, y) &= 8abxy.
\end{aligned}$$

Putting this together, the Hessian matrix at $(0, 0)$ is given by

$$H(0, 0) = \begin{pmatrix} 2 - 4ah & 0 \\ 0 & 2 - 4bh \end{pmatrix}.$$

Since $0 < a < b$, when $h > \frac{1}{2a}$ both eigenvalues are negative, and the critical point $(x, y) = (0, 0)$ represents a local *maximum* of ρ_h. Similarly, when $h < \frac{1}{2b}$ both eigenvalues are positive and $(0, 0)$ is a local *minimum* of ρ_h. When $\frac{1}{2b} < h < \frac{1}{2a}$, one eigenvalue is negative while the other is positive, so $(0, 0)$ is a *saddle point* of ρ_h. Notice that when $h = \frac{1}{2a}$ or $h = \frac{1}{2b}$, one of the eigenvalues is zero and so Theorem 1 does not completely determine the nature of the critical point.

Before proving the theorem, we recall from linear algebra that if H is a symmetric $n \times n$ matrix, then for every vector $\mathbf{u}$,

(1) $\|\mathbf{u}\|^2 \cdot (\text{smallest eigenvalue of } \mathbf{u}) \;\le\; \mathbf{u}^T H \mathbf{u} \;\le\; \|\mathbf{u}\|^2 \cdot (\text{largest eigenvalue of } \mathbf{u}).$

This follows from the Spectral Theorem for symmetric matrices, as will be recalled below.

Proof. Assume that f has a local min at $\mathbf{a}$. This means that there exists $\varepsilon > 0$ such that

$$f(\mathbf{x}) \ge f(\mathbf{a}) \quad \text{whenever} \quad \|\mathbf{x} - \mathbf{a}\| < \varepsilon. \tag{2}$$

Fix any unit vector $\mathbf{u}$, and let $g(t) := f(\mathbf{a} + t\mathbf{u})$.

For $|t| < \varepsilon$, since $\mathbf{u}$ is a unit vector, $\|(\mathbf{a} + t\mathbf{u}) - \mathbf{a}\| = \|t\mathbf{u}\| = |t|\,\|\mathbf{u}\| = |t| < \varepsilon$. So it follows from (2) that $g(t) = f(\mathbf{a} + t\mathbf{u}) \ge f(\mathbf{a}) = g(0)$ whenever $|t| < \varepsilon$. That is, the single-variable function g has a local min at $t = 0$.

1. Now, if f is C^1 then the chain rule, or facts about directional derivatives, imply that g is differentiable, and that

$$0 = g'(0) = \partial_{\mathbf{u}} f(\mathbf{a}) = \mathbf{u} \cdot \nabla f(\mathbf{a}).$$

Since $\mathbf{u}$ was an arbitrary unit vector, it follows that

$$\mathbf{u} \cdot \nabla f(\mathbf{a}) = 0 \qquad \text{for all unit vectors } \mathbf{u}.$$

And this implies that $\nabla f(\mathbf{a}) = 0$. (Make sure that you can justify this statement quickly and efficiently!)

2. Next, continuing to assume that f has a local min at $\mathbf{a}$, if f is C^2, then $g(t)$ as defined above still has a local min at 0, and the chain rule now implies that g is C^2. So we know from first-year calculus that

$$g''(0) \geq 0.$$

Applying the chain rule twice (as we have done several times), we find that

$$g''(0) = \sum_{i,j=1}^{n} u_i u_j \partial_{ij} f(\mathbf{a}) = \mathbf{u}^T H(\mathbf{a})\mathbf{u}.$$

You should do the above computation of g'' until you are thoroughly familiar with it. It follows as above that

$$\mathbf{u}^T H(\mathbf{a})\mathbf{u} \geq 0 \qquad \text{for all unit vectors } \mathbf{u}. \tag{3}$$

In particular, let $\mathbf{u}_i$ be an eigenvector of $H(\mathbf{a})$ (normalized to have length 1), with eigenvalue λ_i. That is, $\mathbf{u}_i$ satisfies

$$H(\mathbf{a})\mathbf{u}_i = \lambda_i \mathbf{u}_i.$$

where $\lambda_i \in \mathbb{R}$ is the associated eigenvalue. Then

$$\mathbf{u}_i^T H(\mathbf{a})\mathbf{u}_i = \lambda_i \mathbf{u}_i^T \mathbf{u}_i = \lambda_i.$$

It follows from this and (3) that $\lambda_i \geq 0$ for all i.

3. We will prove this for $f \in C^3$, although in fact C^2 is enough. Supppose that f is C^3 and that

$$\nabla f(\mathbf{a}) = 0, \qquad \text{all eigenvalues of } H(\mathbf{a}) > 0.$$

Consider $f(\mathbf{a}+\mathbf{h})$ where $\mathbf{h}$ is small. Then according to Taylor's Theorem,

$$f(\mathbf{a}+\mathbf{h}) = f(\mathbf{a}) + \mathbf{h} \cdot \nabla f(\mathbf{a}) + \frac{1}{2}\mathbf{h}^T H(\mathbf{a})\mathbf{h} + r_{2,\mathbf{a}}(\mathbf{a}+\mathbf{h}),$$

Let λ_1 denote the smallest eigenvalue of $H(\mathbf{a})$. Then since $\nabla f(\mathbf{a}) = 0$, and using (1), we see that

$$\begin{aligned} f(\mathbf{a}+\mathbf{h}) &= f(\mathbf{a}) + \frac{1}{2}\mathbf{h}^T H(\mathbf{a})\mathbf{h} + r_{2,\mathbf{a}}(\mathbf{a}+\mathbf{h}) \\ &\geq f(\mathbf{a}) + \|\mathbf{h}\|^2 \left(\frac{\lambda_1}{2} + \frac{r_{2,\mathbf{a}}(\mathbf{a}+\mathbf{h})}{\|\mathbf{h}\|^2} \right) \\ &\geq f(\mathbf{a}) + \|\mathbf{h}\|^2 \left(\frac{\lambda_1}{2} - \frac{|r_{2,\mathbf{a}}(\mathbf{a}+\mathbf{h})|}{\|\mathbf{h}\|^2} \right) \end{aligned}$$

By assumption, $\lambda_1 > 0$, and since f is C^3, we know that[1]

$$\lim_{\mathbf{h}\to 0} \frac{r_{2,\mathbf{a}}(\mathbf{a}+\mathbf{h})}{\|\mathbf{h}\|^2} = 0.$$

[1] The discussion of Taylor series in the online notes shows that if f is C^{k+1}, then the remainder term $r_{k,\mathbf{a}}(\mathbf{a}+\mathbf{h})$ for a kth order Taylor polynomial satisfies $\lim_{\mathbf{h}\to 0} \frac{r_{k,a}(\mathbf{a}+\mathbf{h})}{\|\mathbf{h}\|^k} = 0$. However, this is still true if f is only C^k. This is why it is only necessary for f to be C^2 in this theorem.

It follows that there exists some $\delta > 0$ such that

$$\left| \frac{r_{2,\mathbf{a}}(\mathbf{a}+\mathbf{h})}{\|\mathbf{h}\|^2} \right| \le \frac{\lambda_1}{4} \qquad \text{if } 0 < \|\mathbf{h}\| < \delta.$$

By combining these, we conclude that for such δ,

$$f(\mathbf{a}+\mathbf{h}) \ge f(\mathbf{a}) + \frac{\lambda_1}{4}\|\mathbf{h}\|^2 \qquad \text{if } 0 < \|\mathbf{h}\| < \delta.$$

This proves that $\mathbf{f}$ has a local min at $\mathbf{a}$.

□

Now we present the proof of the linear algebra fact (1) used above. It relies on

Theorem 2 (Spectral Theorem for Symmetric Matrices)**.** *If A is a symmetric $n \times n$ matrix, then it has an orthonormal basis of eigenvectors, and all of its eigenvalues are real numbers.*

That is, there exist real numbers $\lambda_1 \le \lambda_2 \le \cdots \le \lambda_n$, and vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$, such that

$$A\mathbf{v}_i = \lambda_i \mathbf{v}_i \qquad \textit{for } i = 1, \ldots, n \tag{4}$$

and

$$\mathbf{v}_i \cdot \mathbf{v}_j = \begin{cases} 1 & \textit{if } i = j \\ 0 & \textit{if } i \ne j. \end{cases} \tag{5}$$

Now suppose that $\mathbf{u}$ is any vector in $\mathbb{R}^n$. Using the orthonormal basis from the Spectral theorem, we can write $\mathbf{u}$ as a linear combination

$$\mathbf{u} = c_1\mathbf{v}_1 + \cdots + c_n\mathbf{v}_n.$$

Note that

$$\|\mathbf{u}\|^2 = \mathbf{u} \cdot \mathbf{u} = (\sum_{i=1}^n c_i\mathbf{v}_i) \cdot (\sum_{j=1}^n c_j\mathbf{v}_j) = \sum_{i,j=1}^n c_i c_j \mathbf{v}_i \cdot \mathbf{v}_j \overset{(5)}{=} \sum_i^n c_i^2.$$

Also,

$$\mathbf{u}^T H \mathbf{u} = (\sum_{i=1}^n c_i\mathbf{v}_i)^T H (\sum_{j=1}^n c_i\mathbf{v}_j) = \sum_{i,j=1}^n c_i c_j \mathbf{v}_i^T H \mathbf{v}_j$$

And for every i and j,

$$\mathbf{v}_i^T H \mathbf{v}_j \overset{(4)}{=} \mathbf{v}_i^T(\lambda_j \mathbf{v}_j) = \lambda_j \mathbf{v}_i^T \mathbf{v}_j = \lambda_j \mathbf{v}_i \cdot \mathbf{v}_j \overset{(5)}{=} \begin{cases} \lambda_j & \text{if } i = j \\ 0 & \text{if } i \ne j. \end{cases}$$

Inserting this into our calculation above, and recalling that we defined λ_1 to be the smallest eigenvalue, we conclude that

$$\mathbf{u}^T H \mathbf{u} = \sum_{j=1}^n \lambda_j c_j^2 \ge \lambda_1 \sum_{j=1}^n c_j^2 = \lambda_1 \|\mathbf{u}\|^2 = (\text{smallest eigenvalue of } H)\|\mathbf{u}\|^2$$

This is half of fact (1); the other half is established by a small variation of the above argument.

Example 3. It is natural to ask what the geometric significance is of the eigenvalues and eigenvectors of the Hessian matrix. To explore this question, let us first consider the simplest case of curves in the plane and remind ourselves of the definition of curvature. The most intuitive definition of curvature uses the idea of tangent vectors: if we assign to every point on the curve a unit vector in the direction of the tangent line (called a tangent vector at that point), we can think of the curvature of the graph at a point as being given by the amount one needs to rotate the tangent vector there to obtain a tangent vector at a neighbouring point. In Figure 3 below, we see that at points below the x-axis the curvature is much greater than it is at points above the x-axis, as one would expect.

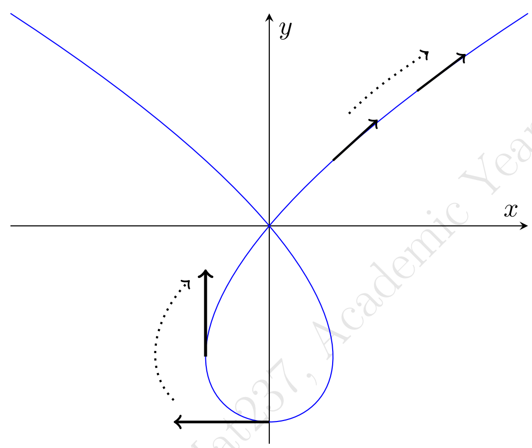


FIGURE 3. Curvature via tangent vectors.

This definition makes it clear that the curvature at a point is related to the rate of change of the slope of the tangent lines near that point. Thus if the curve is given by the graph of a C^2 function, one expects the second derivative to play a role in the computation of curvature. This is indeed the case, the curvature κ_f at the point $(a, f(a))$ on the graph of a function f is given by

$$\kappa_f(a) = \frac{f''(a)}{(1 + f'(a)^2)^{3/2}}.$$

Notice in particular that at critical points of f, the curvature is determined completely by $f''(a)$. Let us consider the more general case of a C^2 function $f : \mathbb{R}^n \to \mathbb{R}$. The problem now is that given a point $\mathbf{a} \in \mathbb{R}^n$, there are many directions tangent to the point $(\mathbf{a}, f(\mathbf{a}))$ on the graph of f. Thus we must talk about the curvature of the graph of f *in a particular direction.* Now suppose $\mathbf{a}$ is a critical point of f, and fix an orthonormal basis of eigenvectors $\{\mathbf{v}_1, ..., \mathbf{v}_n\}$ of $H(\mathbf{a})$ as in Theorem 2. Let us determine the curvature of f in the direction $\mathbf{v}_i$. Define the function $g_i(t) = f(\mathbf{a} + t\mathbf{v}_i)$. Since $\mathbf{a}$ is a critical point of f, we have

$$g_i'(0) = \mathbf{v_i} \cdot \nabla f(\mathbf{a}) = 0.$$

Thus the curvature of f at $\mathbf{a}$ in the direction $\mathbf{v}_i$ is given by

$$\kappa_{g_i}(0) = g_i''(0) = \mathbf{v}_i^T H(\mathbf{a})\mathbf{v}_i = \lambda_i.$$

We have thus obtained a geometric description of the eigenvalues and eigenvectors of $H(\mathbf{a})$! Comparing with (1), we see that ~~$\mathbf{v}_1$ points in the direction of sm~~allest curvature of f, with curvature λ_1. Similarly, $\mathbf{v}_n$ points in the direction of largest curvature, with curvature λ_n. For this reason, when we think of the graph of f as defining a hypersurface in $\mathbb{R}^{n+1}$, the eigenvalues of $H(\mathbf{a})$ are often called the *principal curvatures* of the hypersurface at $\mathbf{a}$, while the eigenvectors $\mathbf{v}_i$ are often called the *principal directions.* When $n = 2$, the determinant $\det H(\mathbf{a}) = \lambda_1\lambda_2$ is called the *Gaussian curvature* at $\mathbf{a}$ of the surface defined by f.

We end these notes with a worked example covering all the material we have discussed.

Example 4. Let us consider the critical points of the surface $z = xy(1 - x^2 - y^2)$ pictured in Figure 4 below.

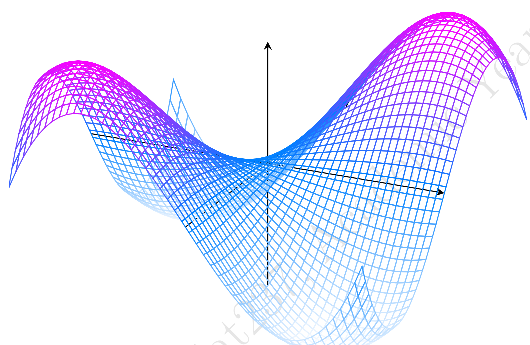


FIGURE 4. The surface $z = xy(1 - x^2 - y^2)$

The partial derivatives of the function $f(x, y) = xy(1 - x^2 - y^2)$ are given by

$$\partial_x f(x,y) = y(1 - 3x^2 - y^2),$$
$$\partial_y f(x,y) = x(1 - x^2 - 3y^2).$$

Solving $\nabla f(x, y) = 0$ shows that the critical points are given by $(0, 0)$, $(0, \pm\frac{1}{\sqrt{3}})$, $(\pm\frac{1}{\sqrt{3}}, 0)$, $(\pm\frac{1}{2}, \pm\frac{1}{2})$. Due to the symmetry of the surface, we will consider the nature of one of each "type" of critical point listed above, the rest can be analyzed similarly. Computing the second partial derivatives of f gives

$$\partial_{xx} f(x,y) = -6xy,$$
$$\partial_{yy} f(x,y) = -6xy,$$
$$\partial_{xy} f(x,y) = 1 - 3x^2 - 3y^2.$$

Plugging in the critical point $(x, y) = (0, 0)$ gives

$$H(0,0) = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

This matrix has eigenvalues $\lambda_1 = -1$, $\lambda_2 = 1$ with corresponding eigenvectors $\mathbf{v}_1 = \frac{1}{\sqrt{2}}(1, -1)^T$, $\mathbf{v}_2 = \frac{1}{\sqrt{2}}(1, 1)^T$. This shows that the origin is a saddle point of f, and the surface $z = f(x, y)$ has largest curvature 1 in the direction of $\frac{1}{\sqrt{2}}(1, 1)^T$, and smallest curvature -1 in the direction of $\frac{1}{\sqrt{2}}(1,-1)^T$. The Gaussian curvature at the origin is -1.

Plugging in the critical point $(0, \frac{1}{\sqrt{3}})$ gives

$$H(0, \frac{1}{\sqrt{3}}) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

Hence the point $(0, \frac{1}{\sqrt{3}})$ is a degenerate critical point of f. The graph of f suggests that it is neither a local maximum nor a local minimum. The curvature at this point is zero in all directions.

Finally, we will check the point $(\frac{1}{2}, \frac{1}{2})$. The formulas above give

$$H(\frac{1}{2}, \frac{1}{2}) = \begin{pmatrix} -\frac{3}{2} & -\frac{1}{2} \\ -\frac{1}{2} & -\frac{3}{2} \end{pmatrix}.$$

This matrix has eigenvalues $\lambda_1 = -2$, $\lambda_2 = -1$ with corresponding eigenvectors $\mathbf{v}_1 = \frac{1}{\sqrt{2}}(1,-1)^T$, $\mathbf{v}_2 = \frac{1}{\sqrt{2}}(1,1)^T$. We conclude that $(\frac{1}{2}, \frac{1}{2})$ is a local maximum for f, and that the curvature at this point is -2 in the direction of $\frac{1}{\sqrt{2}}(1,-1)^T$, and -1 in the direction of $\frac{1}{\sqrt{2}}(1,-1)^T$. The Gaussian curvature at this point is 2.