1. (a) Plugging $f(x,t) = A\sin(x-ct) + B\sin(x+ct)$ into the left side of the wave equation yields

$$\frac{\partial^2 f}{\partial t^2} - c^2\frac{\partial^2 f}{\partial x^2} = A(\sin(x-ct))_{tt} + B(\sin(x+ct))_{tt} - c^2(A(\sin(x-ct))_{xx} + B(\sin(x+ct))_{xx}),$$

so we need to compute these second partials.

We compute that $(\sin(x-ct))_x = \cos(x-ct)$, so $(\sin(x-ct))_{xx} = -\sin(x-ct)$, and similarly $(\sin(x+ct))_{xx} = -\sin(x+ct)$. On the other hand $(\sin(x-ct))_t = -c\cos(x-ct)$, so $(\sin(x-ct))_{tt} = -c^2\sin(x-ct)$, and similarly $(\sin(x+ct))_{tt} = -c^2\sin(x+ct)$ (the sign arising from the derivative of $\cos$ being $-\sin$).

Plugging these into our expression for the left side of the wave equation yields

$$A(-c^2\sin(x-ct)) + B(-c^2\sin(x+ct)) - c^2(-A\sin(x-ct) - B\sin(x+ct)).$$

Combining terms for a common trigonometric function, this can be written as

$$(-Ac^2 + c^2A)\sin(x-ct) + (-Bc^2 + c^2B)\sin(x+ct)$$

that in turn vanishes (by cancellation in the coefficients).

(b) Similarly to (a), we compute that if we plug $f(x,t) = Ah(x-ct) + Bh(x+ct)$ into the left side of the wave equation then we get

$$\frac{\partial^2 f}{\partial t^2} - c^2\frac{\partial^2 f}{\partial x^2} = A(h(x-ct))_{tt} + B(h(x+ct))_{tt} - c^2(A(h(x-ct))_{xx} + B(h(x+ct))_{xx}),$$

so we need to compute these second partials.

Similarly to (a) we obtain $(h(x\pm ct))_{xx} = h''(x\pm ct)$ and $(h(x\pm ct))_{tt} = c^2h''(x\pm ct)$. (For $h = \sin$ we have $h'' = -\cos$.) Hence, our expression for the left side of the wave equation becomes

$$A(c^2h''(x-ct)) + B(c^2h''(x+ct)) - c^2(Ah''(x-ct) + Bh''(x+ct)),$$

and combining terms yields

$$(Ac^2 - c^2A)h''(x-ct) + (Bc^2 - c^2B)h''(x+ct),$$

which vanishes for the same cancellation reason as in (a).

(c) By the identity $\cos(x\pm y) = \cos(x)\cos(y) \mp \sin(x)\sin(y)$, we have

$$\cos(x\pm ct) = \cos x\cos(ct) \mp \sin x\sin(ct).$$

Hence, $\frac{1}{2}\cos(x-ct) - \frac{1}{2}\cos(x+ct)$ is equal to

$$\frac{1}{2}(\cos(x)\cos(ct) + \sin(x)\sin(ct)) - \frac{1}{2}(\cos(x)\cos(ct) - \sin(x)\sin(ct)),$$

and in this expression the cos terms cancel out whereas the sin terms combine to yield $\sin(x)\sin(ct)$ as desired.

$\diamond$

2. (a) We differentiate: $f_t = -(\lambda^2+\mu^2)e^{-(\lambda^2+\mu^2)t}\sin(\lambda x)\sin(\mu y)$, which is just $-(\lambda^2+\mu^2)$ times f itself, i.e. $f_t = -(\lambda^2+\mu^2)f$. Similarly

$$f_x = \lambda e^{-(\lambda^2+\mu^2)t}\cos(\lambda x)\sin(\mu y),\quad f_y = \lambda e^{-(\lambda^2+\mu^2)t}\sin(\lambda x)\cos(\mu y),$$

so

$$f_{xx} = -\lambda^2 e^{-(\lambda^2+\mu^2)t}\sin(\lambda x)\sin(\mu y) = -\lambda^2 f,\quad f_{yy} = -\mu^2 e^{-(\lambda^2+\mu^2)t}\sin(\lambda x)\sin(\mu y) = -\mu^2 f.$$

Thus,

$$f_t = -(\lambda^2+\mu^2)f = -\lambda^2 f - \mu^2 f = f_{xx} + f_{yy}.$$

(b) Since sums and scalar multiplication pass through partial derivatives,

$$\begin{aligned}(a_1f_1+a_2f_2+\cdots+a_Nf_N)_t &= a_1(f_1)_t+a_2(f_2)_t+\cdots+a_N(f_N)_t\\ &= a_1((f_1)_{xx}+(f_1)_{yy})+a_2((f_2)_{xx}+(f_2)_{yy})+\cdots+a_N((f_N)_{xx}+(f_N)_{yy})\\ &= a_1(f_1)_{xx}+a_1(f_1)_{yy}+a_2(f_2)_{xx}+a_2(f_2)_{yy}+\cdots+a_N(f_N)_{xx}+a_N(f_N)_{yy}\end{aligned}$$

We now collect the xx-terms and the yy-terms separately from each other, to arrive at

$$(a_1(f_1)_{xx}+a_2(f_2)_{xx}+\cdots+a_N(f_N)_{xx})+(a_1(f_1)_{yy}+a_2(f_2)_{yy}+\cdots+a_N(f_N)_{yy}),$$

and then pass the sums and scalar multiplications back outside the derivatives to obtain

$$(a_1f_1+a_2f_2+\cdots+a_Nf_N)_{xx}+(a_1f_1+a_2f_2+\cdots+a_Nf_N)_{yy}.$$

Comparing this final expression to the initial t-derivative, we have shown that $a_1f_1+\cdots+a_Nf_N$ satisfies the heat equation, as desired.

(c) For any $k=1,\ldots,N$ the function $e^{-k^2\pi^2t/L}\sin(k\pi x/L)$ satisfies both the initial condition that it becomes $\sin(k\pi x/L)$ when $t=0$ as desired (since the exponential term become 1 when $t=0$), as well as the boundary conditions that it vanishes at $x=0,L$ since the factor $\sin(k\pi x/L)$ becomes $\sin(0)=0$ and $\sin(k\pi)=0$. To check that this function also satisfies the heat equation for each such k, we do a direct calculation of partial derivatives by exploiting that the function has the form $h(t)g(x)$ (so there is no intervention of a product rule for the type of derivatives that we need to compute):

$$(e^{-k^2\pi^2t/L}\sin(k\pi x/L))_t=(e^{-k^2\pi^2t/L})_t\sin(k\pi x/L)=-(k^2\pi^2/L^2)e^{-k^2\pi^2t/L}\sin(k\pi x/L),$$

and

$$\begin{aligned}(e^{-k^2\pi^2t/L}\sin(k\pi x/L))_{xx} &= e^{-k^2\pi^2t/L}(\sin(k\pi x/L))_{xx}\\ &= (k\pi/L)e^{-k^2\pi^2t/L}(\cos(k\pi x/L))_x\\ &= -(k^2\pi^2/L^2)e^{-k^2\pi^2t/L}\sin(k\pi x/L).\end{aligned}$$

By inspection, the two final expressions obtained coincide, so the heat equation holds for each such function for $k=1,2,\ldots,N$.

Now exactly as in (b), if $f_1,\ldots,f_N$ are solutions to the (1-dimensional) heat equation then so is any function of the form $A_1f_1+\cdots+A_Nf_N$ for scalars $A_1,\ldots,A_N$. Using this property and the preceding calculations for each $k=1,\ldots,N$, functions f of the desired general form are always a solution to the heat equation. Likewise, such functions satisfy the desired initial condition and boundary conditions because those are obtained as a linear combination of the initial condition and boundary conditions that have been verified for $e^{-k^2\pi^2t/L}\sin(k\pi x/L)$ for each $k=1,2,\ldots,N$.

◇

3. (a) Using the single-variable Chain Rule in each variable separately, we have

$$f_x=3e^{3x-2y},\ \ f_y=-2e^{3x-2y}.$$

From these expressions we obtain the second partial derivatives:

$$f_{xx}=9e^{3x-2y},\ \ f_{xy}=-6e^{3x-2y},\ \ f_{yy}=4e^{3x-2y}.$$

Putting it all together, we have

$$(\nabla f)(x,y)=\begin{bmatrix}3e^{3x-2y}\\ -2e^{3x-2y}\end{bmatrix},\ \ (\mathrm{H}f)(x,y)=\begin{bmatrix}9e^{3x-2y} & -6e^{3x-2y}\\ -6e^{3x-2y} & 4e^{3x-2y}\end{bmatrix}.$$

(b) Since $f(2,3)=e^0=1$, $(\nabla f)(2,3)=\begin{bmatrix}3\\ -2\end{bmatrix}$, and $(\mathrm{H}f)(2,3)=\begin{bmatrix}9 & -6\\ -6 & 4\end{bmatrix}$, the quadratic approximation at $(2,3)$ is

$$\begin{aligned}f(2+h,3+k) &\approx f(2,3)+\begin{bmatrix}3\\ -2\end{bmatrix}\cdot\begin{bmatrix}h\\ k\end{bmatrix}+\frac{1}{2}\begin{bmatrix}h & k\end{bmatrix}\begin{bmatrix}9 & -6\\ -6 & 4\end{bmatrix}\begin{bmatrix}h\\ k\end{bmatrix}\\ &= 1+(3h-2k)+(1/2)(9h^2-12hk+4k^2)\\ &= 1+(3h-2k)+(9/2)h^2-6hk+2k^2.\end{aligned}$$

(c) Plugging $(h, k) = (0.2, -0.1)$ into the approximation in (b) yields

$$f(2.2, 2.9) \approx 1 + 0.6 + 0.2 + (9/2)(0.04) + 0.12 + 0.02 = 1.8 + 9(0.02) + 0.14 = 1.8 + 0.18 + 0.14 = 2.12.$$

If we had omitted the Hessian term then the approximation would have been 1.8. On a calculator, the "exact" answer is $e^{3(2.2)-2(2.9)} = e^{0.8} \approx 2.2255\dots$. Including the Hessian improves the accuracy.

$\diamond$

4. (a) The associated symmetric 2×2 matrix is $A = \begin{bmatrix} -1 & 6 \\ 6 & -1 \end{bmatrix}$. This has trace -2 and determinant $1 - 36 = -35$, so the characteristic polynomial of A is $\lambda^2 + 2\lambda - 35 = (\lambda - 5)(\lambda + 7)$. Hence, the eigenvalues are $\lambda_1 = 5$ and $\lambda_2 = -7$. To find corresponding eigenvectors we compute

$$A - \lambda_1 \mathrm{I}_2 = A - 5\mathrm{I}_2 = \begin{bmatrix} -6 & 6 \\ 6 & -6 \end{bmatrix}, \quad A - \lambda_2 \mathrm{I}_2 = A + 7\mathrm{I}_2 = \begin{bmatrix} 6 & 6 \\ 6 & 6 \end{bmatrix}$$

whose respective null spaces are spanned by $\mathbf{w}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\mathbf{w}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$.

Since the eigenvalues have opposite signs, $q(x, y)$ is indefinite. Hence, the level sets $q(x, y) = c$ ($c \neq 0$) are hyperbolas, with asymptotes "closer" to the line spanned by $\mathbf{w}_1$ since $|\lambda_1| < |\lambda_2|$ (i.e., $5 < 7$). The hyperbolas crossing the line spanned by $\mathbf{w}_1$ are $q(x, y) = c$ with $c > 0$ since $q(\mathbf{w}_1) = q(1, 1) = 10 > 0$, and the hyperbolas crossing the line spanned by $\mathbf{w}_2$ are $q(x, y) = c$ with $c < 0$ since $q(\mathbf{w}_2) = q(1, -1) = -14 < 0$. The precise picture is as follows (but only the qualitative aspects are expected):

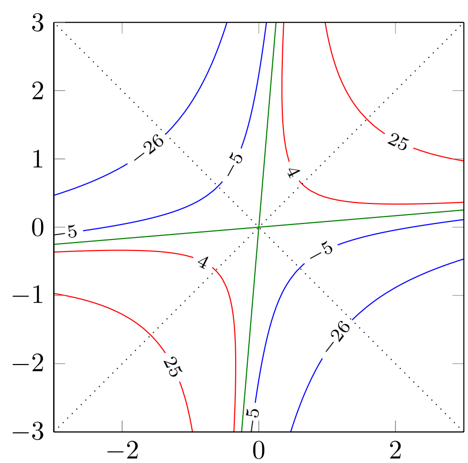

Figure 1: Contour plot for $q(x, y) = -x^2 + 12xy - y^2$ with dotted eigenlines and green asymptotes, and $q(x, y) = c$ in red for $c > 0$ and in blue for $c < 0$.

(b) The associated symmetric 2×2 matrix is $A = \begin{bmatrix} 12 & -3 \\ -3 & 4 \end{bmatrix}$. This has trace 16 and determinant $48 - 9 = 39$, so the characteristic polynomial of A is $\lambda^2 - 16\lambda + 39 = (\lambda - 3)(\lambda - 13)$. Hence, the eigenvalues are $\lambda_1 = 3$ and $\lambda_2 = 13$. To find corresponding eigenvectors we compute

$$A - \lambda_1 \mathrm{I}_2 = A - 3\mathrm{I}_2 = \begin{bmatrix} 9 & -3 \\ -3 & 1 \end{bmatrix}, \quad A - \lambda_2 \mathrm{I}_2 = A - 13\mathrm{I}_2 = \begin{bmatrix} -1 & -3 \\ -3 & -9 \end{bmatrix}$$

whose respective null spaces are spanned by $\mathbf{w}_1 = \begin{bmatrix} 1 \\ 3 \end{bmatrix}$ and $\mathbf{w}_2 = \begin{bmatrix} 3 \\ -1 \end{bmatrix}$.

Since the eigenvalues are both positive, $q(x, y)$ is positive-definite. Hence, the level sets $q(x, y) = c$ (with $c > 0$) are ellipses. These ellipses are longer along the line spanned by $\mathbf{w}_1$ since $|\lambda_1| < |\lambda_2|$ (i.e., $3 < 13$). The precise picture is as follows (but only the qualitative aspects are expected):

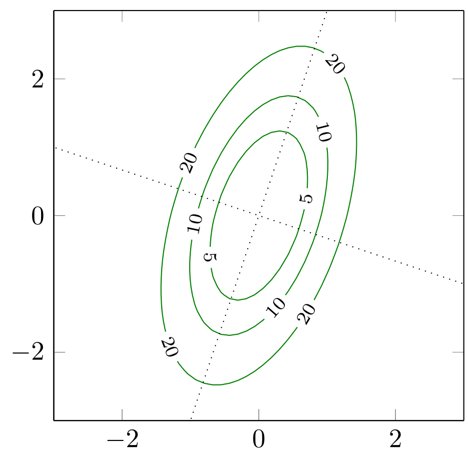

Figure 2: Contour plot for $q(x, y) = 12x^2 - 6xy + 4y^2$ with dotted eigenlines; the ellipses are longer along the direction of the 3-eigenline

◇

5. (a) We have that $f_x = -17x + 4y + 5$, $f_y = 4x - 2y + 2$. Critical points are the solutions of the simultaneous system of linear equations

$$-17x + 4y = -5, \quad 4x - 2y = -2.$$

The unique common solution to this pair of equations is $(x, y) = (1, 3)$, so that is the unique critical point.

(b) The Hessian of f is

$$(\mathrm{H}f)(x, y) = \begin{bmatrix} -17 & 4 \\ 4 & -2 \end{bmatrix}.$$

It is "constant": independent of the point (x, y) at which it is evaluated. The characteristic polynomial of this matrix is $\lambda^2 + 19\lambda + 18$ which factors as $(\lambda + 18)(\lambda + 1)$, so the eigenvalues are $\lambda_1 = -1$ and $\lambda_2 = -18$. These are both negative, so the critical point is a local maximum.

For the eigenvalue $\lambda_1 = -1$ an eigenvector is a nonzero solution to the pair of equations $-17x + 4y = -x$ and $4x - 2y = -y$, which is to say the pair of equations $-16x + 4y = 0$ and $4x - y = 0$. These are both the same line $y = 4x$, so $\mathbf{w}_1 = \begin{bmatrix} 1 \\ 4 \end{bmatrix}$ is such an eigenvector. Likewise, for the eigenvalue $\lambda_2 = -18$, an eigenvector is a nonzero solution to the pair of equations $-17x + 4y = -18x$ and $4x - 2y = -18y$, which is to say the pair of equations $x + 4y = 0$ and $4x + 16y = 0$; these are both the line $x = -4y$ (or equivalently $y = -(1/4)x$) with slope $-1/4$, so $\mathbf{w}_2 = \begin{bmatrix} -4 \\ 1 \end{bmatrix}$ is such an eigenvector.

(c) We shall work with the line $y = 4x$ through $\mathbf{w}_1$ and the line $y = -(1/4)x$ through $\mathbf{w}_2$, translated to cross at the critical point $(1, 3)$. These are lines with respective slopes 4 and $-1/4$ passing through the critical point, the first of these being the t_1-axis and the second being the t_2-axis for coordinates (t_1, t_2) arising from writing everything in terms of $\{\mathbf{w}_1, \mathbf{w}_2\}$, in terms of which the level sets of contour plot near the critical point look approximately like $\lambda_1(\mathbf{w}_1 \cdot \mathbf{w}_1){t_1}^2 + \lambda_2(\mathbf{w}_2 \cdot \mathbf{w}_2){t_2}^2 = c$, which is to say $17(-{t_1}^2 - 18{t_2}^2) = c$ with $c \neq 0$. The coefficients have the same sign, namely negative, so the level curves (for $c < 0$) are ellipses aligned with those two lines as their axes of symmetry.

The contour plot consists of approximate ellipses centered at the critical point $(1, 3)$ aligned along the perpendicular lines with slopes 4 and $-1/4$ (in the directions of the eigenvectors) through the critical point. The longer axis is along the direction of the line through the eigenvector $\mathbf{w}_1$ whose eigenvalue has smaller absolute value. (The approximate ellipses are of the form $-t_1^2 - 18t_2^2 = c$ with $c < 0$, shifted to be centered at $(1, 3)$ and rotated to be aligned with the eigenvector directions, but such explicitness doesn't matter for your answer.)

A detailed picture of the contour plot near the critical point is given in Figure 3 below. You're not expected to come up with anything as precise as this, just to recognize from the Hessian as above that the contour plot consists of nested ellipses around the critical point with approximate axis directions determined by the eigenvalues and eigenvector directions of the Hessian as discussed above and that the longer axis is along the direction of the line through $\mathbf{w}_1$.

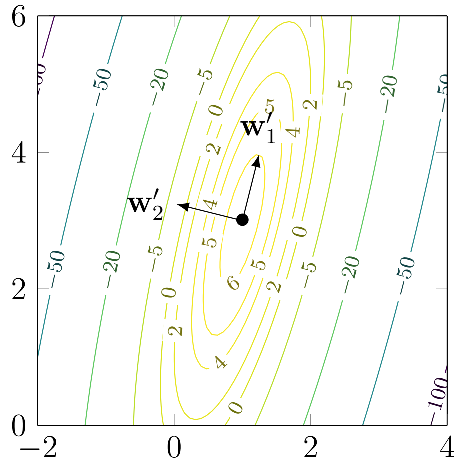

Figure 3: Contour plot of $f(x,y) = -\frac{17}{2}x^2 + 4xy - y^2 + 5x + 2y + 1$ near the critical point (1,3). The vectors $\mathbf{w}'_1$ and $\mathbf{w}'_2$ are unit eigenvectors for the Hessian $(\mathrm{H}f)(1,3)$.

$\diamond$

6. (a) We computed in Exercise I.2.3 that $(\mathrm{H}f)(x,y) = \begin{bmatrix} 16+24x & 6+6y \\ 6+6y & 6x \end{bmatrix}$. Thus $(\mathrm{H}f)(0,0) = \begin{bmatrix} 16 & 6 \\ 6 & 0 \end{bmatrix}$, $(\mathrm{H}f)(0,-2) = \begin{bmatrix} 16 & -6 \\ -6 & 0 \end{bmatrix}$, $(\mathrm{H}f)(-3/2,-1) = \begin{bmatrix} -20 & 0 \\ 0 & -9 \end{bmatrix}$, and $(\mathrm{H}f)(1/6,-1) = \begin{bmatrix} 20 & 0 \\ 0 & 1 \end{bmatrix}$.

   Since a diagonal matrix has its diagonal entries as its eigenvalues, we see that $(-3/2,-1)$ is a local maximum (both eigenvalues are negative) and $(1/6,-1)$ is a local minimum (both eigenvalues are positive). For $(0,0)$, note that the determinant is $(16)(0)-(6)(6) = -36$, so the eigenvalues are nonzero with opposite signs and so this critical point is a saddle point. Finally, at $(0,-2)$, the determinant is $(16)(0)-(-6)(-6) = -36$ again, so once more the critical point is a saddle point.

   (b) The Hessians at $(-3/2,-1)$ and $(1/6,-1)$ are diagonal, so eigenvectors for each are $\mathbf{w}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\mathbf{w}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. Being definite for each, the level sets are approximately ellipses with axes along the coordinate directions. (More specifically, these are respectively approximated by $-20{t_1}^2 - 9{t_2}^2 = c$ with $c < 0$, shifted to be centered at $(-3/2,-1)$, and $20{t_1}^2 + {t_2}^2 = c$ with $c > 0$, shifted to be centered at $(1/6,-1)$; such explicitness doesn't matter for your answer.)

   In both cases the ellipse is somewhat longer vertically than horizontally since in both cases the eigenvalue corresponding to the vertical eigenvector is smaller in absolute value ($|-9| < |-20|$ and $|1| < |20|$).

   A very detailed picture of the contour plot near both critical points is given in Figure 4 below (the green vertical line through $x = 0$ expresses that x is a factor of $f(x,y)$). You're not expected to come up with anything as precise as this, just to recognize from the Hessians as above that the contour plots consist of approximate nested ellipses around the critical points with approximate axis directions determined by the eigenvalues and eigenvector directions of the Hessian as discussed above and that the longer axis along the direction of the line through $\mathbf{w}_2$ in each case. The "zooming in" around $(-1/6,-1)$, which we provide as a bonus, illustrates that the scale on which the quadratic approximation kicks in could be very tiny!

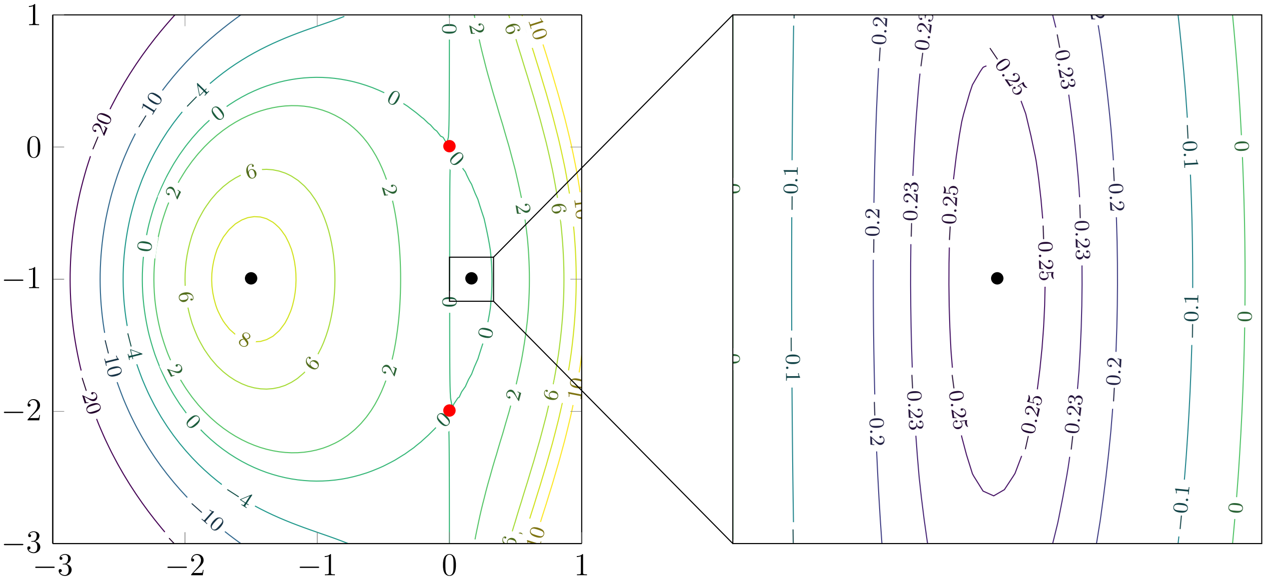

Figure 4: Contour plot of $f(x,y) = 8x^2 + 6xy + 4x^3 + 3xy^2$, with approximate nested ellipses around the critical points $(-3/2,-1)$ and $(-1/6,-1)$ (zoomed in)

$\diamond$

7. (a) The gradient is

$$\nabla f = \begin{bmatrix} \cos x - \sin(x+y) \\ \cos y - \sin(x+y) \end{bmatrix},$$

so the critical points are where $\cos x - \sin(x+y) = 0$ and $\cos y - \sin(x+y) = 0$, or in other words

$$\cos x = \cos y = \sin(x+y).$$

From $\cos x = \cos y$, we have $y = x + 2\pi k$ or $y = -x + 2\pi k$ for an integer k. Since $0 \le x, y < 2\pi$, the first option only happens when $y = x$. Likewise, since $-2\pi < -x \le 0$ and $0 \le y < 2\pi$, the second option only happens when $y = -x + 2\pi$ (or $y = 0 = x$, which is subsumed by the case $y = x$).

First suppose $y = x$, so the condition to be a critical point is $\cos(x) = \sin(2x) = 2\sin(x)\cos(x)$. This says $\cos(x)(1 - 2\sin(x)) = 0$, which is to say $\cos(x) = 0$ or $\sin(x) = 1/2$. By inspecting the unit circle or the graphs of sine and cosine, the case $\cos(x) = 0$ yields the critical points $(\pi/2, \pi/2)$ and $(3\pi/2, 3\pi/2)$ and the case $\sin(x) = 1/2$ yields the critical points $(\pi/6, \pi/6)$ and $(5\pi/6, 5\pi/6)$.

Next suppose $y = -x + 2\pi$, so the condition to be a critical point is $\cos(x) = \sin(y + x) = \sin(2\pi) = 0$, which says $x = \pi/2$ (so $y = 3\pi/2$) or $x = 3\pi/2$ (so $y = \pi/2$). Hence, we get two more critical points: $(\pi/2, 3\pi/2)$, and $(3\pi/2, \pi/2)$

(b) The Hessian of f is

$$(\mathrm{H}f)(x,y) = \begin{bmatrix} -\sin x - \cos(x+y) & -\cos(x+y) \\ -\cos(x+y) & -\sin y - \cos(x+y) \end{bmatrix}.$$

We will compute this at each of the 6 critical points, and work out the definiteness properties to determine the nature of the critical points.

At $(\pi/2, \pi/2)$ we have

$$(\mathrm{H}f)(\pi/2, \pi/2) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

with determinant is $(0)(0) - (1)(1) = -1 < 0$, so this Hessian is indefinite and hence $(\pi/2, \pi/2)$ is a saddle point.

At $(3\pi/2, 3\pi/2)$ we have

$$(\mathrm{H}f)(3\pi/2, 3\pi/2) = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

with determinant $(2)(2)-(1)(1)=3>0$ and trace $2+2=4>0$, so this Hessian is positive-definite and hence $(3\pi/2, 3\pi/2))$ is a local minimum.

At $(\pi/6, \pi/6)$ we have

$$(\mathrm{H}f)(\pi/6, \pi/6) = \begin{bmatrix} -1 & -1/2 \\ -1/2 & -1 \end{bmatrix}$$

with determinant $(-1)(-1)-(1/2)(-1/2)=3/4>0$ and trace $-1-1=-2<0$, so this Hessian is negative-definite and hence $(\pi/6, \pi/6))$ is a local maximum.

At $(5\pi/6, 5\pi/6)$ we have

$$(\mathrm{H}f)(5\pi/6, 5\pi/6) = \begin{bmatrix} -1 & -1/2 \\ -1/2 & -1 \end{bmatrix}$$

with determinant $(-1)(-1)-(1/2)(-1/2)=3/4>0$ and trace $-1-1=-2<0$, so this Hessian is negative-definite and hence $(5\pi/6, 5\pi/6))$ is a local maximum.

At $(\pi/2, 3\pi/2)$ we have

$$(\mathrm{H}f)(\pi/2, 3\pi/2) = \begin{bmatrix} -2 & -1 \\ -1 & 0 \end{bmatrix}$$

with determinant $(-2)(0)-(-1)(-1)=-1<0$, so this Hessian is indefinite and hence $(\pi/2, 3\pi/2)$ is a saddle point.

At $(3\pi/2, \pi/2)$ we have

$$(\mathrm{H}f)(3\pi/2, \pi/2) = \begin{bmatrix} 0 & -1 \\ -1 & -2 \end{bmatrix}$$

with determinant $(0)(-2)-(-1)(-1)=-1<0$, so this Hessian is indefinite and hence $(3\pi/2, \pi/2)$ is a saddle point.

$\diamond$

8. (a) We compute the first and second derivatives:

$$f_x = 3(x^2-y^2),\ f_y = -6xy, \qquad (\mathrm{H}f)(x,y) = \begin{bmatrix} 6x & -6y \\ -6y & -6x \end{bmatrix}.$$

From $f_x=0$ we get $x=\pm y$, and then from $f_y=0$ we see that necessarily $x=y=0$. This is the only critical point, and clearly $(\mathrm{H}f)(0,0)$ is the zero matrix. Thus the second derivative test is completely inconclusive. There are a couple of different approaches to see what happens.

Method 1. Consider the restriction of f to various lines $y=ax$ where one lets a take on different values. We have that $f(x,ax)=x^3-3x(ax)^2=(1-3a^2)x^3$. We see that whenever $1-3a^2\neq 0$, this restriction takes on positive values when x is of one sign (e.g., $x>0$ if $1-3a^2>0$) and takes on negative values when x is of the other sign. Hence, this is an inflection point.

Method 2. Consider the restriction of f to a circle of very small radius r centered at the origin. A way to trace out this circle is via the function $\mathbf{c}(t)=(r\cos t, r\sin t)$ with a choice of small positive r. The restriction of f to this circle corresponds to the composition

$$(f\circ\mathbf{c})(t) = (r\cos t)^3 - 3(r\cos t)(r\sin t)^2 = r^3(\cos^3 t - 3\cos t\sin^2 t) = r^3\cos t(\cos^2 t - 3\sin^2 t).$$

Now we can simply try different values of t. When $t=0$ the value is r^3, while if $t=\pi/4$ then since $\sin\pi/4 = \cos\pi/4 = 1/\sqrt{2}$ the value of f is $r^3(1/\sqrt{2})(1/2-3/2) = -r^3/\sqrt{2}$, which is negative.

By either method, this critical point must be an inflection point since f takes on both positive and negative values arbitrarily near to critical point $(0,0)$. Amusingly, the graph of this function goes by the name "the monkey saddle" because the picture as in Figure 5 shows that it includes room for the tail!

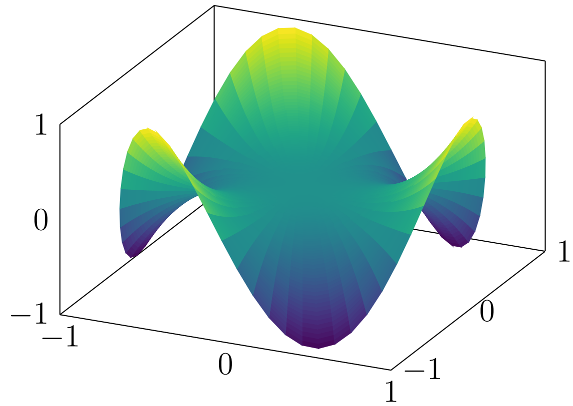

Figure 5: The graph of $x^3 - 3xy^2$ over the unit disk, a so-called "monkey saddle".

(b) For this function we have

$$f_x = 3x^3 + 3x^2,\ f_y = 4y^3,\ f_z = 4z^3, \quad \text{and} \qquad (\mathrm{H}f)(x,y,z) = \begin{bmatrix} 9x^2+6x & 0 & 0 \\ 0 & 12y^2 & 0 \\ 0 & 0 & 12z^2 \end{bmatrix}.$$

Again we compute that the gradient vanishes precisely when $y = z = 0$ and either $x = 0$ or else $x = -1$. Thus there are two critical points, $(0,0,0)$ and $(-1,0,0)$. Clearly the Hessian is not invertible in both cases.

Case 1. For $(0,0,0)$ we define the function of one variable $h(x) = f(x,0,0) = \frac{3}{4}x^4 + x^3$, and we look at the behavior of this function as x near 0 changes sign. The idea here is that x^4 is much smaller than x^3 when x is near 0, so the sign of x^3 is the same as the sign of $\frac{3}{4}x^4 + x^3$ for x near 0, and this changes sign – it is positive when $x > 0$ and negative when $x < 0$ with $|x|$ small. Therefore this critical point is an inflection point.

Case 2. For $(-1,0,0)$ let us try the same method again, namely consider the same function $h(x) = \frac{3}{4}x^4 + x^3$ for x now near -1, so $f(x,y,z) = h(x) + y^4 + z^4$. This satisfies $h(-1) = \frac{3}{4} - 1 = -\frac{1}{4}$, $h'(x) = 3x^3 + 3x^2$ so $h'(-1) = 0$, $h''(x) = 9x^2 + 6x$ so $h''(-1) = 3$. This shows that the Taylor expansion of $h(x)$ near $x = -1$ is

$$h(x) \approx -\frac{1}{4} + \frac{3}{2}(x+1)^2$$

to second order in $x+1$ near 0. The second-order term $(3/2)(x+1)^2$ is nonnegative (and positive when $x \neq -1$), and the other part of f, $y^4 + z^4$, is also nonnegative. We conclude that $(-1,0,0)$ is a local minimum for f (in fact, it is also a global minimum).

(c) We compute

$$f_x = 3x^2 - 4x + y^2,\ f_y = 2xy, \qquad (\mathrm{H}f)(x,y) = \begin{bmatrix} 6x-4 & 2y \\ 2y & 2x \end{bmatrix}.$$

The condition $f_y = 0$ says $x = 0$ or $y = 0$. The further condition $f_x = 0$ then gives that if $x = 0$ then $y = 0$, whereas if $y = 0$ then $x(3x-4) = 0$.

Therefore, there are two critical points: $(0,0)$ and $(4/3,0)$. The Hessian of f at $(0,0)$ and $(4/3,0)$ are

$$(\mathrm{H}f)(0,0) = \begin{bmatrix} -4 & 0 \\ 0 & 0 \end{bmatrix},\ (\mathrm{H}f)(4/3,0) = \begin{bmatrix} 4 & 0 \\ 0 & 8/3 \end{bmatrix}.$$

Thus, $(4/3,0)$ is a local minimum because the Hessian of f at that point is positive-definite.

To analyze what is happening near $(0,0)$, which fails the second derivative test, we extract the common factor of x in all terms to rewrite the function in a more convenient form:

$$f(x,y) = x(x^2 - 2x + y^2) = x(x^2 - 2x + 1 + y^2 - 1) = x((x-1)^2 + y^2 - 1).$$

The part inside parentheses on the right vanishes on the circle $(x-1)^2 + y^2 = 1$, is negative for points (x,y) inside this circle, and is positive for points outside that circle. Notice that this is a circle of radius 1 with center $(1,0)$, so it passes through the origin. Multiplying this function by x maintains the signs of this function in the region $x > 0$ but makes this function negative for $x < 0$.

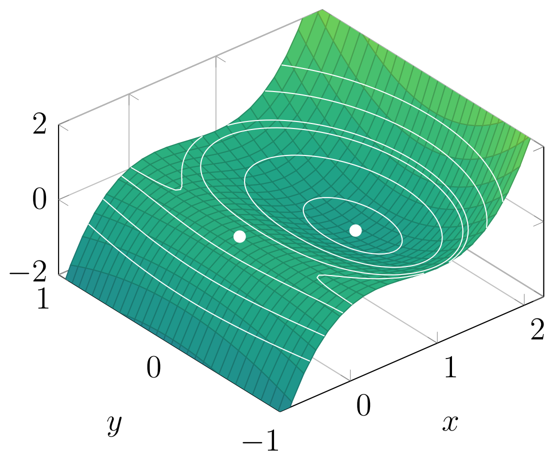

Figure 6: The graph of $x^3 - 2x^2 + xy^2$ with the two critical points $(0, 0)$ and $(4/3, 0)$ shown. The surface has a valley with bottom at $(4/3, 0)$, and a slight inflection at $(0, 0)$.

From all of this information we see that f is negative both when $x < 0$ and when $x > 0$ with (x, y) inside that circle, but f is positive when $x > 0$ and (x, y) is outside that circle. There are points of this latter sort as close to the origin as we wish, so $(0, 0)$ (at which the value of f is 0) is an inflection point.

The surface graph of f in Figure 6 illustrates these features, with the white curves indicating level sets: the positive level curves approach the white dot over the origin in a pinched hyperbolic shape over $x = 0$ (the y-axis), whereas the negative level curves approach that white dot in approximately straight curves that we see both coming from the valley side and from the opposite side (consider walking on the surface over the negative x-axis up towards the white dot over the origin).

◇

9. The Hessian of f at (a, b) is

$$(\mathrm{H}f)(a,b) = \begin{bmatrix} f_{xx}(a,b) & f_{xy}(a,b) \\ f_{xy}(a,b) & f_{yy}(a,b) \end{bmatrix}.$$

Laplace's equation says that the diagonal entries are negatives of each other, so the trace of $(\mathrm{H}f)(a, b)$ vanishes. Furthermore, $\det(\mathrm{H}f)(a, b) = f_{xx}(a, b)f_{yy}(a, b) - f_{xy}(a, b)^2 = -(f_{xx}(a, b)^2 + f_{xy}(a, b)^2) \le 0$ (since the negative of the square of a function is always ≤ 0).

These facts together imply that the eigenvalues λ_1, λ_2 of $(\mathrm{H}f)(a, b)$ satisfy $\lambda_2 = -\lambda_1$ and $\lambda_1\lambda_2 \le 0$. This means that either both are zero, or else one of them is > 0 and and the other is < 0. If the eigenvalues both vanish then the Hessian is the zero matrix (since the associated quadratic form vanishes), and otherwise it is indefinite; in the latter case (a, b) is a saddle point for f. ◇

10. This function has

$$\nabla f(x,y) = \begin{bmatrix} y \\ x \end{bmatrix}, \quad \text{and} \qquad (\mathrm{H}f)(x,y) = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

So the general formula for gradient descent tells us that with input $\begin{bmatrix} p_1 \\ p_2 \end{bmatrix}$ the output is

$$\begin{bmatrix} q_1 \\ q_2 \end{bmatrix} = \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} - (0.25)\begin{bmatrix} p_2 \\ p_1 \end{bmatrix} = \begin{bmatrix} p_1 - (0.25)p_2 \\ p_2 - (0.25)p_1 \end{bmatrix}.$$

For these eight cases we respectively get

$$\begin{bmatrix} 1 \\ -0.25 \end{bmatrix}, \begin{bmatrix} 0.375 \\ 0.375 \end{bmatrix}, \begin{bmatrix} -0.25 \\ 1 \end{bmatrix}, \begin{bmatrix} -0.625 \\ 0.625 \end{bmatrix}, \begin{bmatrix} -1 \\ 0.25 \end{bmatrix}, \begin{bmatrix} -0.375 \\ -0.375 \end{bmatrix}, \begin{bmatrix} 0.25 \\ -1 \end{bmatrix}, \begin{bmatrix} 0.625 \\ -0.625 \end{bmatrix}.$$

You can see in Figure 7 that in all cases other than $\pm(1/2, 1/2)$ the point $g(\mathbf{p})$ is further away from the critical point $(0, 0)$.

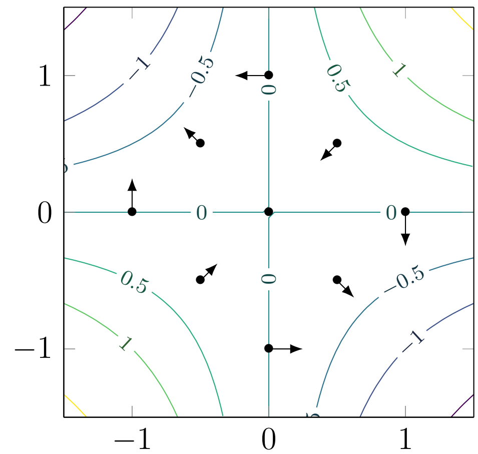

Figure 7: The first step of gradient descent for $f(x, y) = xy$ near the saddle point $(0, 0)$. For each point p, we have drawn an arrow from p to the output $p - (0.25)\nabla f(p)$.

◇

11. (a) As suggested consider the difference

$$k(t) = (6 + 2t) - f(3 - 4t, 4 - 2t).$$

By the Chain Rule, we obtain

$$k'(t) = 2 - (\nabla f)(3 - 4t, 4 - 2t) \cdot \begin{bmatrix} -4 \\ -2 \end{bmatrix};$$

the minus sign in front of the gradient is due to the order of subtraction in the definition of $k(t)$. If you think about the partial derivatives involved then another application of the Chain Rule gives

$$k''(t) = -\begin{bmatrix} -4 \\ -2 \end{bmatrix}^\top \Big((\mathrm{H}f)(3 - 4t, 4 - 2t)\Big) \begin{bmatrix} -4 \\ -2 \end{bmatrix},$$

where $\begin{bmatrix} -4 \\ -2 \end{bmatrix}^\top$ appears due to t-differentiation of the point $(3 - 4t, 4 - 2t)$ at which ∇f is evaluated in the formula for $k'(t)$.

Because of the assumption that $\mathrm{H}f$ is positive definite at *every* point, the minus sign in front of the formula for $k''(t)$ implies that $k''(t) < 0$ for every t.

(b) By design, $k(0) = k(1) = 0$. For any $h : [a, b] \to \mathbf{R}$ with $h(a) = h(b) = 0$ and $h''(x) < 0$ for all $a < x < b$, we want to show $h(x) > 0$ for every $a < x < b$ (and this can then be applied to $k : [0, 1] \to \mathbf{R}$). We will consider the minima of $h(x)$ on $[a, b]$, aiming to show these occur only at the endpoints (where h vanishes, so it must be positive over the interval strictly between those endpoints, as desired).

Suppose x_0 is a point in $[a, b]$ where h reaches its minimum. If x_0 is not one of the endpoints then $h'(x_0) = 0$ (because it is a local extremum away from the endpoints) and $h''(x_0) < 0$ (since h'' is negative on the interval (a, b) strictly between the endpoints by hypothesis). But then by the second derivative test, x_0 would be a local maximum for h, yet we chose it to be a minimum for h on the interval $[a, b]$. If a point is both a minimum and a maximum for a function then the function must be constant, so h would have to be constant, an absurdity since $h'' < 0$ everywhere on (a, b).

We conclude that h attains its minimal values on $[a, b]$ only at the endpoints, where we know it vanishes, so for all $a < x < b$ we have $h(x) > 0$ as we wanted to show.

◇

