Let $M = [a_{ij}]$ be a matrix. The **matrix norm** $\|M\|$ is defined by

$$\|M\| = \max\{|M\mathbf{x}| : |\mathbf{x}| = 1\}.$$

In short, $\|M\|$ is the farthest from the origin (in the image space) to which M maps a point from the unit sphere (in the domain space). This maximum exists because $|M\mathbf{x}|$ is a continuous real-valued function restricted to $|\mathbf{x}| = 1$, a compact domain, and therefore attains a maximum.

$\|M\|$ is different from $\sqrt{\sum_{i,j} a_{ij}^2}$, the norm M would have if you regarded it as a vector (say, by stringing all the entries out in one row). Let us call that second norm $|M|_E$ (for Euclidean norm).

1. Consider the matrix $M = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$. Show that $\|M\|$ and $|M|_E$ are different. (For this simple matrix, both norms are easy to compute without any theory.)

2. Show that, for every $\mathbf{x}$, $|M\mathbf{x}| \le \|M\||\mathbf{x}|$. Thus $M\mathbf{x} = O(\mathbf{x})$ for every matrix M, something we have long asserted, but which we had to waive our hands to prove previously. Now we can apply the definition of O directly.

3. A function f is **uniformly continuous** if

$$\forall\epsilon\ \exists\delta\ \forall\mathbf{a} \quad |\mathbf{x} - \mathbf{a}| < \delta \implies |f(\mathbf{x}) - f(\mathbf{a})| < \epsilon.$$

In other words, you can pick δ solely in terms of ϵ; you don't have to pick different δ's depending on $\mathbf{a}$. (Not every function is uniformly continuous, as you can see by considering $f(x) = x^2$; for any ϵ vertical tolerance, the bigger a is, the steeper the graph is and the narrower the δ tolerance has to be.)

Prove: any linear transformation T is uniformly continuous. Hint: let M be the matrix of T and use the matrix norm.

Note: the definition of regular continuity (for a function continuous everywhere) is

$$\forall\mathbf{a}\ \forall\epsilon\ \exists\delta \quad |\mathbf{x} - \mathbf{a}| < \delta \implies |f(\mathbf{x}) - f(\mathbf{a})| < \epsilon.$$

Notice the difference in order of quantifiers. That difference really makes a difference!

4. Let S be a $n \times n$ symmetric matrix, such as one gets by taking all the second partial derivatives of some continuously differentiable $f : R^n \to R$. Consider the quadratic form $Q(\mathbf{x}) = \mathbf{x}^{\mathrm{T}} S\mathbf{x}$. Then it is a theorem of linear algebra that there is a matrix

$$\Lambda = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix},$$

and for each $\mathbf{x}$ there is another vector $\mathbf{v}$, such that

1. $|\mathbf{x}| = |\mathbf{v}|$, and

2. $Q(\mathbf{x}) = \mathbf{v}^{\mathrm{T}}\Lambda\mathbf{v}$.

(This is because S can be diagonalized using an orthonormal change of basis; $\mathbf{v}$ is just $\mathbf{x}$ rewritten relative to this other basis. The λ_i in Λ are just the eigenvalues of S, which are necessarily real numbers. We may take $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$.)

a) Show that, for all $\mathbf{x} \in R^n$,

$$\lambda_1|\mathbf{x}|^2 \geq Q(\mathbf{x}) \geq \lambda_n|\mathbf{x}|^2,$$

with each extreme obtained for some vectors. Hint: rewrite everything in terms of $\mathbf{v}$ and Λ.

b) Let $c = \max\{|\lambda_1|, |\lambda_n|\}$. Show that $\|S\| = c$.

5. In the previous problem, we found a simple formula for $\|M\|$ in terms of eigenvalues, when M is symmetric. This formula can't possibly work for general matrices M, because M might not even be square, and thus won't have any eigenvalues. Even if M is square, the eigenvalues might not be real. Even if the eigenvalues are real, the formula for $\|M\|$ in the last problem doesn't always hold. Let $M = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$.

a) Show that all the eigenvalues are 1.

b) Find a unit vector $\mathbf{x}$, such that $|M\mathbf{x}| > 1$, thus showing $\|M\| > 1$.

6. The discussions in Edwards, Sect II.4 and II.8, show how to find $\|M\|$ for any matrix using Lagrange multipliers. However, there is a much cleaner way using linear algebra. See if you can prove: for any M,

$$\|M\| = \sqrt{\mu}, \quad \text{where } \mu \text{ is the largest eigenvalue of } M^{\mathrm{T}}M.$$

Hint: rewrite $|M\mathbf{x}|^2 = (M\mathbf{x}) \cdot (M\mathbf{x})$ using matrix multiplication.

7. Return to the matrix $M = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$.

a) Find a unit vector $\mathbf{u}$ such that $|M\mathbf{u}|$ is maximum (for unit vectors).

b) Find a unit vector $\mathbf{v}$ such that $|M\mathbf{v}|$ is minimum. Is there any nice geometric relationship between $\mathbf{u}$ and $\mathbf{v}$?

c) The set of points $\{\mathbf{x} = (x, y) : |M\mathbf{x}| = 1\}$ is an ellipse. Can you find a quadratic equation in x and y for this ellipse? (It won't be as simple as $ax^2 + by^2 = 1$ because the axes of the ellipse aren't the x- and y-axes, but there must be a equation of the form $ax^2+bxy+cy^2 = 1$.

d) The set $\{M\mathbf{x} : |\mathbf{x}| = 1\}$ is an ellipse. Can you find a quadratic equation for this ellipse? What do the lengths of the axes of this ellipse have to do with $\|M\|$?

8. Let $\mathbf{a}$ be a critical point of $f : R^n \to R$, so that

$$f(\mathbf{a}+\mathbf{h}) \;=\; f(\mathbf{a}) + \mathbf{h}^{\mathrm{T}} S\mathbf{h} + R(\mathbf{h}),$$

where S is the symmetric matrix of second partials (the "Hessian"), and $R(\mathbf{h}) \in O(|\mathbf{h}|^3)$. Suppose S is positive definite with minimum eigenvalue λ_m. Let C be the constant such that $R(\mathbf{h}) \leq C|\mathbf{h}|^3$ for $\mathbf{h}$ sufficiently small. In terms of these quantities, find a δ such that

$$0 < |\mathbf{h}| < \delta \implies f(\mathbf{a}+\mathbf{h}) > f(\mathbf{a}).$$

In other words, prove conclusively that if the Hessian is positive definite at a critical point $\mathbf{a}$, then $\mathbf{a}$ is a local minimum of f (and there are no other local minima within δ of $\mathbf{a}$).

9. Determine if the quadratic form

$$Q(x, y, z) = x^2 + 3y^2 + 5z^2 + 4xy + 6xz + 8yz$$

is positive definite, negative definite, or neither. Hint: write Q in the form $\mathbf{x}^{\mathrm{T}} A\mathbf{x}$, where A is a symmetric matrix. The principal determinant method is easy to apply, the eigenvalue method is more tedious, at least if you do it without machine help.