Day 20

1. The chain rule

Recall from single variable calculus that if $y = f(x)$ and $x = u(t)$ are differentiable functions, then the derivative of y with respect to t is given by

$$\frac{dy}{dt} = \frac{dy}{dx}\frac{dx}{dt} = f'(x)u'(t).$$

What can we say when one or both of these functions is replaced by a multivariable function? Let's consider the simplest case first.

THEOREM 12.32 **Chain Rule (Version I)**

Given functions $z = f(x, y)$, $x = u(t)$, and $y = v(t)$, for all values of t at which u and v are differentiable, and if f is differentiable at $(u(t), v(t))$, then

$$\frac{dz}{dt} = \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt}.$$

For example, if $z = x^2 e^{xy}$ and $x = t^2$, $y = \ln t$, then

$$\begin{aligned}\frac{dz}{dt} &= \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt}\\ &= (2xe^{xy} + x^2 y e^{xy}) \cdot 2t + x^3 e^{xy} \cdot \frac{1}{t}\\ &= (2t^2 e^{t^2 \ln t} + t^4 \ln(t) e^{t^2 \ln t})2t + t^6 e^{t^2 \ln t}\frac{1}{t}.\end{aligned}$$

In-class exercise[1]**:** Let $z = x^2 y$ and $x = 3t$, $y = t^2$. Find $\frac{dz}{dt}$ first by using the chain rule and then by substitution and differentiation with respect to t.

Now let's consider the case when x and y are themselves functions of multiple variables:

THEOREM 12.33 **Chain Rule (Version II)**

Given functions $z = f(x, y)$, $x = u(s, t)$, and $y = v(s, t)$, for all values of s and t at which u and v are differentiable, and if f is differentiable at $(u(s, t), v(s, t))$, then

$$\frac{\partial z}{\partial s} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial s} \quad \text{and} \quad \frac{\partial z}{\partial t} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial t}.$$

For example, let $z = \frac{x}{y}$ and $x = s^2$, $y = 1 + s + 2t$. Then

$$\frac{\partial z}{\partial s} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial s} = \frac{1}{y} \cdot 2s + \left(-\frac{x}{y^2}\right) \cdot 1 = \frac{1}{1+s+2t}2s - \frac{s^2}{(1+s+2t)^2}$$

[1]In both cases, it works out to be $\frac{dz}{dt} = 54t^2$.

and

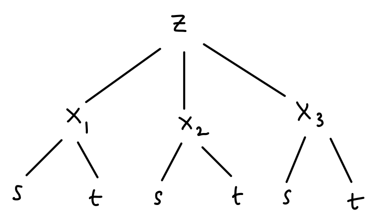
$$\frac{\partial z}{\partial t} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial t} = \frac{1}{y}\cdot 0 + \Big(-\frac{x}{y^2}\Big)\cdot 2 = -2\frac{s^2}{(1+s+2t)^2}.$$

In-class exercise[2]: Let $z = x^2y^3$ and $x = t\sin(s)$, $y = s\cos(t)$. Find $\frac{\partial z}{\partial s}$ and $\frac{\partial z}{\partial t}$.

If we continue to increase the number of variables in our functions, you can guess what the corresponding chain rule will look like. For example, given functions $z = f(x_1, x_2, x_3)$, $x_1 = u_1(s,t)$, $x_2 = u_2(s,t)$, and $x_3 = u_3(s,t)$, the partial derivatives $\frac{\partial z}{\partial s}$ and $\frac{\partial z}{\partial t}$ are given by

$$\frac{\partial z}{\partial s} = \frac{\partial z}{\partial x_1}\frac{\partial x_1}{\partial s} + \frac{\partial z}{\partial x_2}\frac{\partial x_2}{\partial s} + \frac{\partial z}{\partial x_3}\frac{\partial x_3}{\partial s} \qquad \text{and} \qquad \frac{\partial z}{\partial t} = \frac{\partial z}{\partial x_1}\frac{\partial x_1}{\partial t} + \frac{\partial z}{\partial x_2}\frac{\partial x_2}{\partial t} + \frac{\partial z}{\partial x_3}\frac{\partial x_3}{\partial t}.$$

The most general form of the chain rule can be found in Theorem 12.34 in the textbook. I would not advise memorizing it, but rather understand the pattern. It's often helpful to map out the variable dependencies using a tree diagram. For example, we could represent the functions $z = f(x_1, x_2, x_3)$, $x_1 = u_1(s,t)$, $x_2 = u_2(s,t)$, and $x_3 = u_3(s,t)$ as follows:



To find $\frac{\partial z}{\partial s}$, we would:

1. Find the branches that end with s:

$$z \to x_1 \to s, \qquad z \to x_2 \to s, \qquad z \to x_3 \to s$$

2. "Differentiate along" each such branch:

$$\frac{\partial z}{\partial x_1}\frac{\partial x_1}{\partial s}, \qquad \frac{\partial z}{\partial x_2}\frac{\partial x_2}{\partial s}, \qquad \frac{\partial z}{\partial x_3}\frac{\partial x_3}{\partial s}$$

3. Add all terms together:

$$\frac{\partial z}{\partial s} = \frac{\partial z}{\partial x_1}\frac{\partial x_1}{\partial s} + \frac{\partial z}{\partial x_2}\frac{\partial x_2}{\partial s} + \frac{\partial z}{\partial x_3}\frac{\partial x_3}{\partial s}.$$

We would use similar steps to find $\frac{\partial z}{\partial t}$. Note that this technique is very versatile; it works for any number of variables composed in any way.

[2] $\frac{\partial z}{\partial s} = 2t^2s^3\sin(s)\cos(s)\cos(t)^3 + 3t^2s^2\sin(s)^2\cos(t)^3$,
$\frac{\partial z}{\partial t} = 2ts^3\sin(s)^2\cos(t)^3 - 3t^2s^3\sin(s)^2\cos(t)^2\sin(t)$

Day 21

1. Proof of the chain rule

Yesterday we saw various versions of the chain rule, including a general pattern that allows one to find the partial derivatives of any composition of functions. Let's see a sketch of the proof of the simplest version, which states that if $z = f(x, y)$, $x = u(t)$, and $y = v(t)$, then

$$\frac{dz}{dt} = \frac{\partial z}{\partial x}\frac{dx}{dt} + \frac{\partial z}{\partial y}\frac{dy}{dt} = f_x(x, y)u'(t) + f_y(x, y)v'(t).$$

The idea is to use the linear approximation of each of the functions z, x, and y. Fix a point t_0, and let $x_0 = u(t_0)$ and $y_0 = v(t_0)$. Starting with the linear approximation of z, we have

$$f(x, y) - f(x_0, y_0) \approx f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0)$$

(Recall that this follows from the differentiability of f, or equivalently the existence of a tangent plane to the graph of f at the point $(x_0, y_0, f(x_0, y_0))$.) The linear approximations of x and y are given by

$$x - x_0 = u(t) - u(t_0) \approx u'(t_0)(t - t_0) \qquad \text{and} \qquad y - y_0 = v(t) - v(t_0) \approx v'(t_0)(t - t_0).$$

Substituting these into the first approximation and dividing both sides by $t - t_0$, we get

$$\frac{f(x, y) - f(x_0, y_0)}{t - t_0} \approx f_x(x_0, y_0)u'(t_0) + f_y(x_0, y_0)v'(t_0).$$

Finally, taking the limit as $t \to t_0$, the approximation becomes an equality and the left-hand side converges to the derivative of z at $t = t_0$. Thus,

$$\frac{dz}{dt} = f_x(x_0, y_0)u'(t_0) + f_y(x_0, y_0)v'(t_0),$$

which is what we needed to show.

2. The gradient

The *gradient* of a function f is simply the vector containing its partial derivatives. If f is a function of x and y, then its gradient is

$$\nabla f(x, y) = \langle f_x(x, y), f_y(x, y)\rangle.$$

Likewise if f is a function of x, y, and z, then its gradient is

$$\nabla f(x, y, z) = \langle f_x(x, y, z), f_y(x, y, z), f_z(x, y, z)\rangle.$$

For now, we'll treat the gradient as a vector defined at each point of differentiability. However, we can also think of it as a function that takes a vector input ($\langle x, y\rangle$ or $\langle x, y, z\rangle$) and produces a vector output ($\langle f_x(x, y), f_y(x, y)\rangle$ or $\langle f_x(x, y, z), f_y(x, y, z), f_z(x, y, z)\rangle$). We'll adopt that perspective when we study vector fields later in the course.

3. The gradient and directional derivatives

The gradient has several important properties and uses. Firstly, it can be used to calculate directional derivatives:

THEOREM 12.36 **The Gradient and the Directional Derivative**

Let $f(x, y)$ be a function of two variables and (x_0, y_0) be a point in the domain of f at which f is differentiable. Then, for every unit vector $\mathbf{u} \in \mathbb{R}^2$,

$$D_{\mathbf{u}} f(x_0, y_0) = \nabla f(x_0, y_0) \cdot \mathbf{u}.$$

Similarly, if $\mathbf{u} \in \mathbb{R}^3$ is a unit vector, $f(x, y, z)$ is a function of three variables, and (x_0, y_0, z_0) is a point in the domain of f at which f is differentiable, then

$$D_{\mathbf{u}} f(x_0, y_0, z_0) = \nabla f(x_0, y_0, z_0) \cdot \mathbf{u}.$$

To see why this is true, let's again use linear approximation. (A more rigorous proof based on the chain rule can be found in the textbook.) Let $\mathbf{u} = \langle \alpha, \beta \rangle$ and let $h > 0$. Then

$$f(x_0 + \alpha h, y_0 + \beta h) - f(x_0, y_0) \approx f_x(x_0, y_0)\alpha h + f_y(x_0, y_0)\beta h = (\nabla f(x_0, y_0) \cdot \mathbf{u})h.$$

Dividing both sides by h and letting $h \to 0$, we get

$$\nabla f(x_0, y_0) \cdot \mathbf{u} = \lim_{h \to 0} \frac{f(x_0 + \alpha h, y_0 + \beta h) - f(x_0, y_0)}{h} = D_{\mathbf{u}} f(x_0, y_0).$$

Let's revisit the example of $f(x, y) = xy^2$ and $\mathbf{u} = \langle \sqrt{3}/2, 1/2 \rangle$ from last week. Using the definition of the directional derivative, we showed that

$$D_{\mathbf{u}} f(x, y) = xy + \frac{\sqrt{3}}{2} y^2.$$

Now we can get the same result more quickly using the gradient:

$$D_{\mathbf{u}} f(x, y) = \nabla f(x, y) \cdot \mathbf{u} = \langle y^2, 2xy \rangle \cdot \left\langle \frac{\sqrt{3}}{2}, \frac{1}{2} \right\rangle = \frac{\sqrt{3}}{2} y^2 + xy.$$

4. The gradient and direction of fastest increase

Another key property of the gradient of f is that it points in the direction in which f is increasing most rapidly. More precisely, suppose $\nabla f(x_0, y_0) \neq \mathbf{0}$, and let

$$\mathbf{u_0} = \frac{\nabla f(x_0, y_0)}{\|\nabla f(x_0, y_0)\|}$$

be the unit vector in the direction of $\nabla f(x_0, y_0)$. Then

$$D_{\mathbf{u_0}} f(x_0, y_0) \geq D_{\mathbf{u}} f(x_0, y_0)$$

for every unit vector $\mathbf{u}$. (An analogous statement holds for functions of three variables.)

Let's see why this is true. Let $\mathbf{u}$ be an arbitrary unit vector, and let θ be angle between $\mathbf{u}$ and $\nabla f(x_0, y_0)$. Then

$$D_{\mathbf{u}} f(x_0, y_0) = \nabla f(x_0, y_0) \cdot \mathbf{u} = \|\nabla f(x_0, y_0)\| \|\mathbf{u}\| \cos\theta = \|\nabla f(x_0, y_0)\| \cos\theta.$$

The right-hand side is maximized when $\cos\theta = 1$, or $\theta = 0$. This happens if and only if $\mathbf{u} = \mathbf{u_0}$. Therefore $D_{\mathbf{u}}f(x_0, y_0) \leq D_{\mathbf{u_0}}f(x_0, y_0)$.

Remarks:

- While $\nabla f(x_0, y_0)$ tells us the direction of fastest increase, the rate of change in that direction is given by $\|D_{\mathbf{u_0}}f(x_0, y_0)\|$, which turns out to be $\|\nabla f(x_0, y_0)\|$.
- The same arguments show that $-\nabla f(x_0, y_0)$ points in the direction of fastest decrease, and the rate of change in that direction is $-\|\nabla f(x_0, y_0)\|$.

5. The gradient and level curves/surfaces

Recall that if f is a function of two variables, then its level curve at height k is the set of points (x, y) that solve $f(x, y) = k$. Each point (x_0, y_0) in the domain of f belongs to a unique level curve, namely $f(x, y) = f(x_0, y_0)$.

A final important property of the gradient is that $\nabla f(x_0, y_0)$ is always orthogonal to the level curve that passes through (x_0, y_0). More precisely, let $\mathbf{r}(t) = \langle x(t), y(t)\rangle$ be a parametrization of the level curve through (x_0, y_0), and let t_0 be such that $\mathbf{r}(t_0) = \langle x_0, y_0\rangle$. Then

$$\nabla f(x_0, y_0) \cdot \mathbf{r}'(t_0) = 0.$$

(In other words, $\nabla f(x_0, y_0)$ is orthogonal to any tangent vector to the level curve at (x_0, y_0).)

To see why this is true, let $F(t) = f(x(t), y(t))$. Then F is constant, because $f(x, y) = f(x_0, y_0)$ for all (x, y) belonging to the level curve through (x_0, y_0). So $F'(t_0) = 0$. But we can also apply the chain rule to see that

$$F'(t_0) = f_x(x_0, y_0)x'(t_0) + f_y(x_0, y_0)y'(t_0) = \nabla f(x_0, y_0) \cdot \mathbf{r}'(t_0).$$

Equating these two expressions for $F'(t_0)$, we get $\nabla f(x_0, y_0) \cdot \mathbf{r}'(t_0) = 0$.

An analogous statement holds for functions of three variables. Each point (x_0, y_0, z_0) in the domain of a three-variable function f belongs to a unique level surface, given by $f(x, y, z) = f(x_0, y_0, z_0)$. The same argument as above shows that if $\mathbf{r}$ is any vector function taking values within this level surface, and if t_0 is such that $\mathbf{r}(t_0) = \langle x_0, y_0, z_0\rangle$, then

$$\nabla f(x_0, y_0, z_0) \cdot \mathbf{r}'(t_0) = 0.$$

6. Tangent planes revisited

We know how to find the tangent plane to a graph: Given a function $f(x, y)$, the tangent plane to the graph at the point $(x_0, y_0, f(x_0, y_0))$ is given by the equation

$$f_x(x_0, y_0)(x - x_0) + f_y(x_0, y_0)(y - y_0) = z - f(x_0, y_0).$$

However, not every surface in $\mathbb{R}^3$ is a graph. For example, there is no function whose graph is the unit sphere given by $x^2 + y^2 + z^2 = 1$.

Instead, a generic surface in $\mathbb{R}^3$ is given by an equation of the form $f(x, y, y) = k$ and is thus a level surface. Given a point (x_0, y_0, z_0) on this surface, the vector $\nabla f(x_0, y_0, z_0)$ is normal to the surface at that point. Therefore, it's also normal to the tangent plane. This implies the tangent plane has equation

$$\nabla f(x_0, y_0, z_0) \cdot \langle x - x_0, y - y_0, z - z_0\rangle = 0.$$

Day 22

1. Summary of the gradient

Yesterday we saw the definition of the gradient of a function, along with a few of its properties. Specifically, recall that if f is a function of x and y, then its gradient at the point (x, y) is the vector

$$\nabla f(x, y) = \langle f_x(x, y), f_y(x, y)\rangle.$$

We showed that the gradient has the following three important properties:

1. *Directional derivatives.* If $\mathbf{u}$ is any unit vector, then the directional derivative $D_{\mathbf{u}} f(x_0, y_0)$ is given by $D_{\mathbf{u}} f(x_0, y_0) = \nabla f(x_0, y_0) \cdot \mathbf{u}$.
2. *Direction of fastest increase.* The gradient points in the direction in which f is increasing most rapidly. In other words, $D_{\mathbf{u}} f(x_0, y_0)$ is maximal when $\mathbf{u} = \frac{1}{\|\nabla f(x_0,y_0)\|}\nabla f(x_0, y_0)$. (Similarly, $-\nabla f(x_0, y_0)$ points in the direction of fastest decrease.)
3. *Orthogonal to level curves.* Let C be the level curve of f that passes through (x_0, y_0). Then $\nabla f(x_0, y_0)$ is orthogonal to C at (x_0, y_0).

If f is a three-variable function, then very similar statements hold. (In the third statement, replace "level curve" with "level surface".)

In-class exercise: Let $f(x, y, z) = e^{2x-y+z}$.

(a) Find $D_{\mathbf{u}} f(1, 0, 1)$ where $\mathbf{u}$ is the unit vector in the direction of $\langle 1, 2, 3\rangle$.
(b) Find $D_{\mathbf{v}} f(1, 0, 1)$ where $\mathbf{v}$ is the unit vector in the direction of fastest increase of f.
(c) Find an equation for the tangent plane to the level surface $e^{2x-y+z} = e^3$ at the point $(1, 0, 1)$.

Solution:

(a) We can see that $\nabla f(x, y, z) = \langle 2e^{2x-y+z}, -e^{2x-y+z}, e^{2x-y+z}\rangle$, and thus $\nabla f(1, 0, 1) = e^3\langle 2, -1, 1\rangle$. We're given that $\mathbf{u} = \frac{1}{\sqrt{14}}\langle 1, 2, 3\rangle$. Using Property 1, we get

$$D_{\mathbf{u}} f(1, 0, 1) = \nabla f(1, 0, 1) \cdot \mathbf{u} = \frac{3e^3}{\sqrt{14}}.$$

(b) By Property 2, the direction of fastest increase at $(1, 0, 1)$ is given by $\nabla f(1, 0, 1)$. Thus $\mathbf{v} = \frac{1}{\|\nabla f(1,0,1)\|}\nabla f(1, 0, 1) = \frac{1}{\sqrt{6}}\langle 2, -1, 1\rangle$. So

$$D_{\mathbf{v}}(1, 0, 1) = \nabla f(1, 0, 1) \cdot \mathbf{v} = \sqrt{6}.$$

(Note that $\sqrt{6} = \|\nabla f(1, 0, 1)\|$. This is not a coincidence. Try proving that if $\mathbf{w} = \frac{1}{\|\nabla f(x,y,z)\|}\nabla f(x, y, z)$, then $D_{\mathbf{w}} f(x, y, z) = \|\nabla f(x, y, z)\|$.)

(c) By Property 3, we know that the vector $\nabla f(1, 0, 1) = \langle 2, -1, 1\rangle$ is orthogonal to the level surface through $(1, 0, 1)$. This vector is therefore normal to the tangent plane at $(1, 0, 1)$. Since we have a normal vector to the plane and a point on the plane, we can write down an equation for the plane:

$$2(x - 1) - y + z - 1 = 0.$$

It was pointed out in class that the level surface $e^{2x-y+z} = e^3$ already *is* a plane: We can take the natural logarithm of both sides to get $2x - y + z = 3$, the same plane we just found. In general, a level surface will not be a plane, but it was a great observation nevertheless.

Day 23

1. Optimization

Recently we've seen a few uses for the gradient of a multivariable function. Another important application of the gradient is in solving optimization problems; that is, finding the maximum and minimum values of a function. Many problems in applied fields such as engineering or data science can be phrased as optimization problems. For example, essentially all machine learning algorithms are trained by optimizing a function of several variables (often hundreds or thousands). One of the oldest, but still widely used, training algorithms is called *gradient descent*. It's based the principle that a function can be minimized by moving incrementally in the direction opposite the gradient.[1]

To study optimization of multivariable functions, we need some definitions:

Local and Global Extrema of a Function of Two Variables

(a) A function $f(x, y)$ has a ***local*** or ***relative maximum*** at (x_0, y_0) if $f(x_0, y_0) \geq f(x, y)$ for every point (x, y) in some open disk containing (x_0, y_0).

(b) A function $f(x, y)$ has a ***local*** or ***relative minimum*** at (x_0, y_0) if $f(x_0, y_0) \leq f(x, y)$ for every point (x, y) in some open disk containing (x_0, y_0).

(c) A function $f(x, y)$ has a ***local*** or ***relative extremum*** at (x_0, y_0) if f has either a local maximum or a local minimum at (x_0, y_0).

(d) A function $f(x, y)$ has a ***global*** or ***absolute maximum*** at (x_0, y_0) if $f(x_0, y_0) \geq f(x, y)$ for every point (x, y) in the domain of f.

(e) A function $f(x, y)$ has a ***global*** or ***absolute minimum*** at (x_0, y_0) if $f(x_0, y_0) \leq f(x, y)$ for every point (x, y) in the domain of f.

(f) A function $f(x, y)$ has a ***global*** or ***absolute extremum*** at (x_0, y_0) if f has either a global maximum or a global minimum at (x_0, y_0).

Note that the definitions of local and global extrema extend easily to functions of three or more variables. We have two goals in studying optimization: Learn how to find local extrema and how to classify them as local minima or maxima. We may also consider the problem of finding absolute extrema, but this generally more difficult, particularly if the domain of the function is not an open set.

2. Finding local extrema

The key to finding local extrema is the observation that such points can only occur where the gradient is zero or where f fails to be differentiable. If, for a given f, there are not too many points where ∇f is zero or where f is non-differentiable, then we can find the local extrema of f by inspecting each such point. Here are the precise definitions we need:

[1]This doesn't always work, due to the existence of local minima, but it's very useful nevertheless.

Stationary Points of a Function of Two Variables

A point (x_0, y_0) in the domain of a function $f(x, y)$ is called a ***stationary point*** of f if f is differentiable at (x_0, y_0) and $\nabla f(x_0, y_0) = \mathbf{0}$.

Critical Points of a Function of Two Variables

A point (x_0, y_0) in the domain of a function $f(x, y)$ is called a ***critical point*** of f if (x_0, y_0) is a stationary point of f or if f is not differentiable at (x_0, y_0).

And the main theorem:

Local Extrema Occur at Critical Points

If $f(x, y)$ has a local extremum at (x_0, y_0), then (x_0, y_0) is a critical point of f.

A few comments: First, note that the definitions and theorem above can be extended to functions of three or more variables. Second, note that the theorem will be more useful once we're able to decide whether a given critical point is actually a local extremum. (We'll say more about this soon.) But already it's helpful, as we'll see below. Finally, let's think about why the theorem is true: Suppose f has a local extremum, let's say a local maximum, at (x_0, y_0). If f is not differentiable at (x_0, y_0), then (x_0, y_0) is a critical point. So let's assume f is differentiable at (x_0, y_0). If $\nabla f(x_0, y_0) \neq \mathbf{0}$, then $\nabla f(x_0, y_0)$ points in the direction of fastest increase of f; in particular, if we move any (small) distance in the direction of $\nabla f(x_0, y_0)$, we will increase the value of f. But this contradicts the fact that (x_0, y_0) is a local maximum, so we must have $\nabla f(x_0, y_0) = \mathbf{0}$, and thus (x_0, y_0) is a critical point.

As an example, let's solve the following problem: Find the dimensions of the box with volume 1000 and minimal surface area.

Let x, y, and z denote the side-lengths of the box. Since $xyz = 1000$ and the problem is symmetric with respect to each side, we can guess that the answer will be $x = y = z = 10$. But how can we show this more rigorously? Let's first write down a formula for the surface area:

$$A = 2(xy + xz + yz).$$

We need to use the fact that $xyz = 1000$. We can do this by solving for one of the variables, say z, and substituting into the surface area formula:

$$z = \frac{1000}{xy} \qquad \Rightarrow \qquad A = 2\Big(xy + \frac{1000}{y} + \frac{1000}{x}\Big)$$

Thus A is a function of x and y defined on the open set $\{(x, y) \in \mathbb{R}^2 \colon x, y > 0\}$. We want to find the location of its global minimum, which will also be a local minimum (since the domain of A is open). Therefore, by the theorem above, the global minimum will occur at a critical point of A. Since A is differentiable, this critical point will be a stationary point, i.e. a point (x_0, y_0) where $\nabla A(x_0, y_0) = \mathbf{0}$. We can see that

$$\nabla A(x, y) = \Big\langle 2\Big(y - \frac{1000}{x^2}\Big), 2\Big(x - \frac{1000}{y^2}\Big)\Big\rangle.$$

Setting this equal to $\mathbf{0}$ and solving for x and y, we get $x = y = 10$. Thus the global minimum of A occurs at $(10, 10)$. Finally, note that if $x = y = 10$, then we must have $z = 10$ as well, since $xyz = 1000$. So in conclusion, the side-lengths are indeed $x = y = z = 10$ as we predicted.