Lagrange multipliers, examples

Examples of the Lagrangian and Lagrange multiplier technique in action.

Background

- [Introduction to Lagrange multipliers](#)
- [Gradient](#)
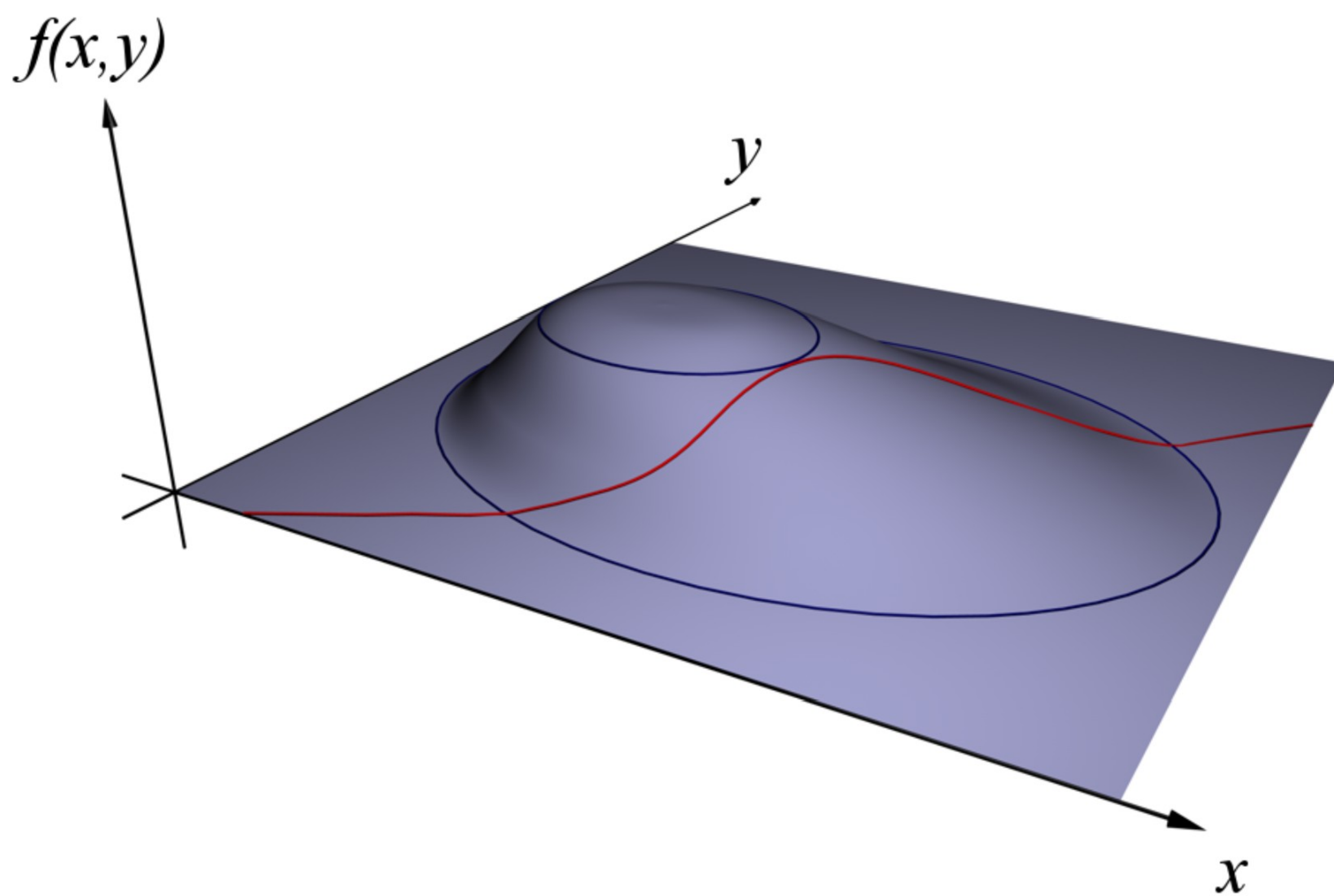
Lagrange multiplier technique, quick recap

When you want to maximize (or minimize) a multivariable function $f(x, y, \dots)$ subject to the constraint that another multivariable function equals a constant, $g(x, y, \dots) = c$, follow these steps:

- **Step 1**: Introduce a new variable λ, and define a new function $\mathcal{L}$ as follows:

$$\mathcal{L}(x, y, \dots, \lambda) = f(x, y, \dots) - \lambda(g(x, y, \dots) - c)$$

  This function $\mathcal{L}$ is called the "Lagrangian", and the new variable λ is referred to as a "Lagrange multiplier"

- **Step 2**: Set the gradient of $\mathcal{L}$ equal to the zero vector.

$$\nabla \mathcal{L}(x, y, \ldots, \lambda) = \mathbf{0} \quad \leftarrow \text{Zero vector}$$

In other words, find the **critical points** of $\mathcal{L}$.

- **Step 3**: Consider each solution, which will look something like $(x_0, y_0, \ldots, \lambda_0)$. Plug each one into f. Or rather, first remove the λ_0 component, then plug it into f, since f does not have λ as an input. Whichever one gives the greatest (or smallest) value is the maximum (or minimum) point your are seeking.

Example 1: Budgetary constraints

Problem

Suppose you are running a factory, producing some sort of widget that requires steel as a raw material. Your costs are predominantly human labor, which is \$20 per hour for your workers, and the steel itself, which runs for \$170 per ton. Suppose your revenue R is loosely modeled by the following equation:

$$R(h, s) = 200h^{2/3}s^{1/3}$$

- h represents hours of labor
- s represents tons of steel

If your budget is \$20,000, what is the maximum possible revenue?

Solution

The \$20 per hour labor costs and \$170 per ton steel costs tell us that the total cost of production, in terms of h and s, is

$$20h + 170s$$

Therefore the budget of \$20,000 can be translated to the constraint

$$20h + 170s = 20{,}000$$

Before we dive into the computation, you can get a feel for this problem using the following interactive diagram. You can see which values of (h, s) yield a given revenue (blue curve) and which values satisfy the constraint (red line).
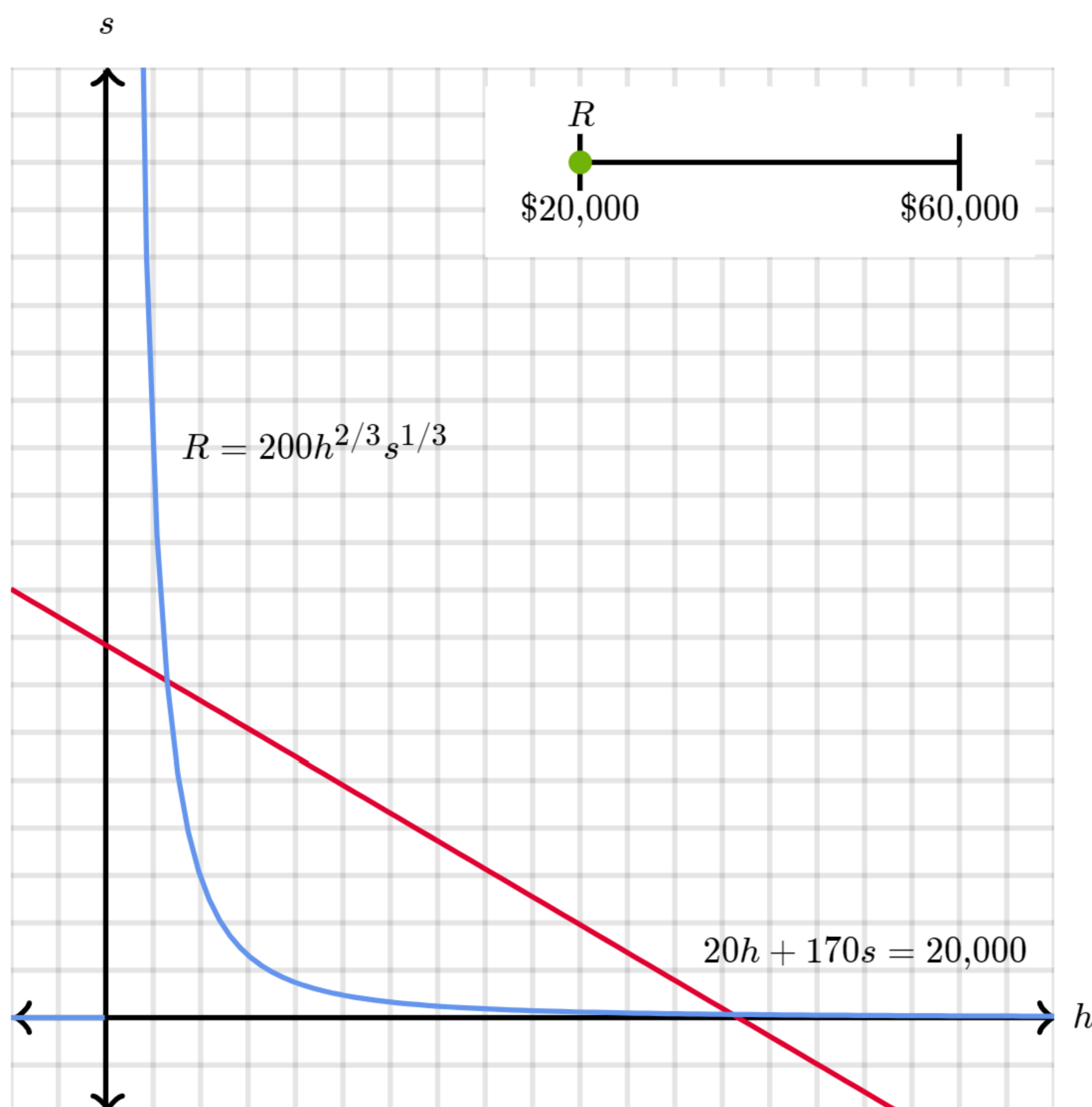
Since we need to maximize a function $R(h, s)$, subject to a constraint, $20h + 170s = 20{,}000$, we begin by writing the Lagrangian function for this setup:

$$\mathcal{L}(h, s, \lambda) = 200h^{2/3}s^{1/3} - \lambda(20h + 170s - 20{,}000)$$

Next, set the gradient $\nabla\mathcal{L}$ equal to the **0** vector. This is the same as setting each partial derivative equal to 0. First, we handle the partial derivative with respect to h.

$$0 = \frac{\partial\mathcal{L}}{\partial h}$$

$$0 = \frac{\partial}{\partial h}\left(200h^{2/3}s^{1/3} - \lambda(20h + 170s - 20{,}000)\right)$$

$$0 = 200 \cdot \frac{2}{3}h^{-1/3}s^{1/3} - 20\lambda$$

Next, we handle the partial derivative with respect to s.

$$0 = \frac{\partial\mathcal{L}}{\partial s}$$

$$0 = \frac{\partial}{\partial s}\left(200h^{2/3}s^{1/3} - \lambda(20h + 170s - 20{,}000)\right)$$

$$0 = 200 \cdot \frac{1}{3}h^{2/3}s^{-2/3} - 170\lambda$$

Finally we set the partial derivative with respect to λ equal to 0, which as always is just the same thing as the constraint. In practice, you can of course just write the constraint itself, but I'll write out the partial derivative here just to make things clear.

$$0 = \frac{\partial \mathcal{L}}{\partial \lambda}$$

$$0 = \frac{\partial}{\partial \lambda}(200h^{2/3}s^{1/3} - \lambda(20h + 170s - 20{,}000))$$

$$0 = -20h - 170s + 20{,}000$$

$$20h + 170s = 20{,}000$$

Putting it together, the system of equations we need to solve is

$$0 = 200 \cdot \frac{2}{3}h^{-1/3}s^{1/3} - 20\lambda$$

$$0 = 200 \cdot \frac{1}{3}h^{2/3}s^{-2/3} - 170\lambda$$

$$20h + 170s = 20{,}000$$

In practice, you should almost always use a computer once you get to a system of equations like this. Especially because the equation will likely be more complicated than these in real applications. Once you do, you'll find that the answer is

$$h = \frac{2{,}000}{3} \approx 666.667$$

$$s = \frac{2{,}000}{51} \approx 39.2157$$

$$\lambda = \sqrt[3]{\frac{8{,}000}{459}} \approx 2.593$$

This means you should employ about 667 hours of labor, and purchase 39 tons of steel, which will give a maximum revenue of

$$R(667, 39) = 200(667)^{2/3}(39)^{1/3} \approx \boxed{\$51{,}777}$$

The interpretation of this constant $\lambda = 2.593$ is left to the [next article](#)

Example 2: Maximizing dot product

Problem: Let the three-dimensional vector $\vec{\mathbf{v}}$ be defined as follows.

$$\vec{\mathbf{v}} = \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix}$$

Consider every possible unit vector $\hat{\mathbf{u}}$ in three-dimensional space. For which one is the dot product $\hat{\mathbf{u}} \cdot \vec{\mathbf{v}}$ the greatest?

The diagram below is two-dimensional, but not much changes in the intuition as we move to three dimensions.
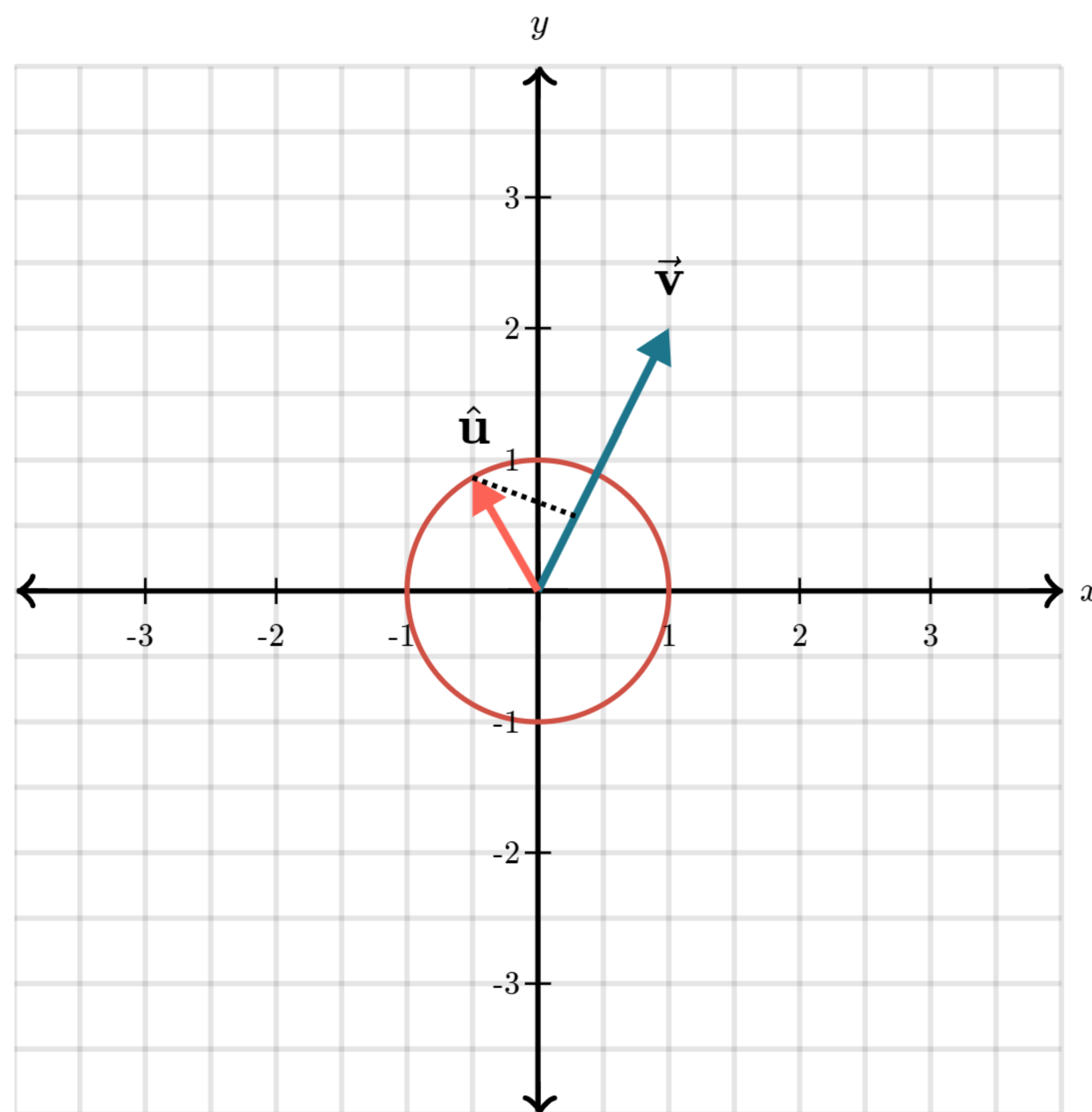
Two-dimensional analogy to the three-dimensional problem we have. Which unit vector $\hat{\mathbf{u}}$ maximizes the dot product $\hat{\mathbf{u}} \cdot \vec{\mathbf{v}}$?

If you are fluent with dot products, you may already know the answer. It's one of those mathematical facts worth remembering. If you don't know the answer, all the better! Because we will now find and prove the result using the Lagrange multiplier method.

Solution:

First, we need to spell out how exactly this is a constrained optimization problem. Write the coordinates of our unit vectors as x, y and z:

$$\hat{\mathbf{u}} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$$

The fact that $\hat{\mathbf{u}}$ is a **unit vector** means its magnitude is 1:

$$||\hat{\mathbf{u}}|| = \sqrt{x^2 + y^2 + z^2} = 1$$

$$\Downarrow$$

$$x^2 + y^2 + z^2 = 1$$

This is our constraint.

Maximizing $\hat{\mathbf{u}} \cdot \vec{v}$ means maximizing the following quantity:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix} = 2x + 3y + z$$

The Lagrangian, with respect to this function and the constraint above, is

$$\mathcal{L}(x, y, z, \lambda) = 2x + 3y + z - \lambda(x^2 + y^2 + z^2 - 1).$$

We now solve for $\nabla \mathcal{L} = \mathbf{0}$ by setting each partial derivative of this expression equal to 0.

$$\frac{\partial}{\partial x}(2x + 3y + z - \lambda(x^2 + y^2 + z^2 - 1)) = 2 - \lambda 2x = 0$$

$$\frac{\partial}{\partial y}(2x + 3y + z - \lambda(x^2 + y^2 + z^2 - 1)) = 3 - \lambda 2y = 0$$

$$\frac{\partial}{\partial z}(2x + 3y + z - \lambda(x^2 + y^2 + z^2 - 1)) = 1 - \lambda 2z = 0$$

Remember, setting the partial derivative with respect to λ equal to 0 just restates the constraint.

$$\frac{\partial}{\partial \lambda}(2x + 3y + z - \lambda(x^2 + y^2 + z^2 - 1)) = -x^2 - y^2 - z^2 + 1 = 0$$

Solving for x, y and z in the first three equations above, we get

$$x = 2 \cdot \frac{1}{2\lambda}$$

$$y = 3 \cdot \frac{1}{2\lambda}$$

$$z = 1 \cdot \frac{1}{2\lambda}$$

Ah, what beautiful symmetry. Each of these expressions has the same $\frac{1}{2\lambda}$ factor, and the coefficients 2, 3 and 1 match up with the coordinates of $\vec{v}$. Being good math students as we are, we won't let good symmetry go to waste. In this case, combining the three equations above into a single vector equation, we can relate $\hat{u}$ and $\vec{v}$ as follows:

$$\hat{u} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \frac{1}{2\lambda} \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix} = \frac{1}{2\lambda} \vec{v}$$

Therefore $\hat{u}$ is proportional to $\vec{v}$! Geometrically, this means $\hat{u}$ points in the same direction as $\vec{v}$. There are two unit vectors proportional $\vec{v}$,
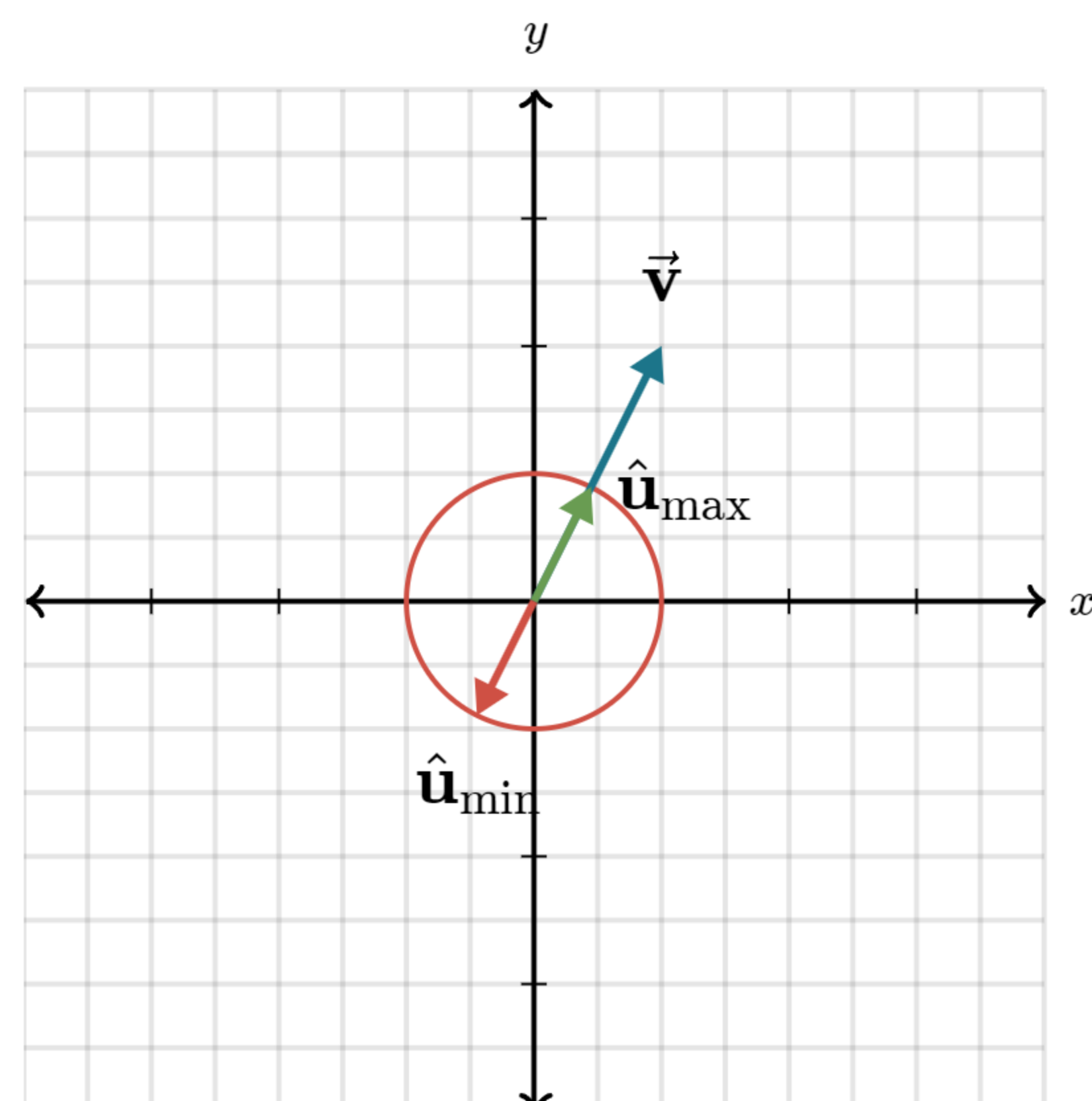
- One which points in the same direction, this is the vector that maximizes $\hat{u} \cdot \vec{v}$.
- One which points in the opposite direction. This one minimizes $\hat{u} \cdot \vec{v}$.

Two-dimensional analogy showing the two unit vectors which maximize and minimize the quantity $\hat{u} \cdot \vec{v}$.

We can write these two unit vectors by normalizing $\vec{v}$, which just means dividing $\vec{v}$ by its magnitude:

$$\hat{u}_{\max} = \frac{\vec{v}}{||\vec{v}||}$$

$$\hat{u}_{\min} = -\frac{\vec{v}}{||\vec{v}||}$$

The magnitude $||\vec{v}||$ is $\sqrt{2^2 + 3^2 + 1^2} = \sqrt{14}$, so we can write the maximizing unit vector $\hat{u}_{\max}$ explicitly as like this:

$$\hat{\mathbf{u}}_{\text{max}} = \begin{bmatrix} 2/\sqrt{14} \\ 3/\sqrt{14} \\ 1/\sqrt{14} \end{bmatrix}$$

Just skip the Lagrangian

If you read the [last article](), you'll recall that the whole point of the Lagrangian $\mathcal{L}$ is that setting $\nabla \mathcal{L} = 0$ encodes the two properties a constrained maximum must satisfy:

- Gradient alignment between the target function and the constraint function,

$$\nabla f(x, y) = \lambda \nabla g(x, y)$$

- The constraint itself,

$$g(x, y) = c$$

When working through examples, you might wonder why we bother writing out the Lagrangian at all. Wouldn't it be easier to just start with these two equations rather than re-establishing them from $\nabla \mathcal{L} = 0$ every time? The short answer is yes, it would be easier. If you find yourself solving a constrained optimization problem by hand, and you remember the idea of gradient alignment, feel free to go for it without worrying about the Lagrangian.

In practice, it's often a computer solving these problems, not a human. Given that there are many highly optimized programs for finding when the gradient of a given function is 0, it's both clean and useful to encapsulate our problem into the equation $\nabla \mathcal{L} = 0$.

Furthermore, the Lagrangian itself, as well as several functions deriving from it, arise frequently in the theoretical study of optimization. In this light, reasoning about the single object $\mathcal{L}$ rather than multiple conditions makes it easier to see the connection between high-level ideas. Not to mention, it's quicker to write down on a blackboard.

In either case, whatever your future relationship with constrained optimization might be, it is good to be able to think about the Lagrangian itself and what it does. The examples above illustrate how it works, and hopefully help to drive home the point that $\nabla \mathcal{L} = 0$ encapsulates both $\nabla f = \lambda \nabla g$ and $g(x, y) = c$ in a single equation.