This is the most amazing tool. The gradient of a function will always be orthogonal to the level curve of that function at any given point. To see why, take the first order Taylor expansion around a given point. For perturbations around the point to remain on the level curve (i.e. g(x) = g(x+dx)), the perturbation direction must be orthogonal to the gradient. Therefore, the gradient of g(x) will always be orthogonal to the tangential direction of the curve g(x) = c.

# Open Questions

This section contains discussion around open questions I have

## Interpretation of min/max operations

The repo containing the code for this post is located [here]( https://github.com/chrisnielsen/chrisnielsen.github.io)

A question I have struggled with that is how to interpret the expression:

\[\mathop {\max }\limits\_x \mathop {\min }\limits\_y f\left( {x,y} \right)\]

So far, I think the best way to interpret this expression is the following: consider breaking the expression into two components as follows:

\[\mathop {\max }\limits\_x \mathop {\min }\limits\_y f\left( {x,y} \right) = \mathop {\max }\limits\_x g\left( x \right)\] where \[g\left( x \right) = \mathop {\min }\limits\_y f\left( {x,y} \right)\]

### What can we do to fix this

Therefore, we can interpret this expression as:

1. Pick an x value that we are evaluating
2. Evaluate g(x) for that x value (this will return the y value that minimizes f(x,y) for that choice of x)
3. Repeat step 1 until we find the value of x which maximizes g(x)

Another way to interpret the min/max operations is to think about finding the saddle point of a function. Finding the saddle point is equivalent to finding the min/max solution for a set of variables.

This is a test.

This is a test.

\[\begin{align}

& \log p\left( x;\theta \right)=\log \int{p\left( x,z;\theta \right)\,}dz \\

& \,\,\,\,\,\,\,\,\,\,\,\,\,\,\,\,\,\,\,\,\,\,\,\,=\sqrt{{{b}^{2}}-4ac} \\

& \,\,\,\,\,\,\,\,\,\,\,\,\,\,\,\,\,\,\,\,\,\,\,\,=\log \int{p\left( x,z;\theta \right)\,}dz \\

& \,\,\,\,\,\,\,\,\,\,\,\,\,\,\,\,\,\,\,\,\,\,\,\,=\log \int{p\left( x,z;\theta \right)\,}dz \\

& \,\,\,\,\,\,\,\,\,\,\,\,\,\,\,\,\,\,\,\,\,\,\,\,=\sqrt{{{b}^{2}}-4ac} \\

& \,\,\,\,\,\,\,\,\,\,\,\,\,\,\,\,\,\,\,\,\,\,\,\,=\log \int{p\left( x,z;\theta \right)\,}dz \\

\end{align}\]

### Here is a proof that for any function\[f\left( {x,y} \right)\]we have \[\mathop {\max }\limits\_y \mathop {\min }\limits\_x f\left( {x,y} \right) \le \mathop {\min }\limits\_x \mathop {\max }\limits\_y f\left( {x,y} \right)\]

Proof

Suppose that \[\left( {{x^\*},{y^\*}} \right)\]is the optimal solution to \[\mathop {\min }\limits\_x \mathop {\max }\limits\_y f\left( {x,y} \right)\], such that \[\mathop {\min }\limits\_x \mathop {\max }\limits\_y f\left( {x,y} \right) = f\left( {{x^\*},{y^\*}} \right)\]. This implies that \[\mathop {\min }\limits\_x \mathop {\max }\limits\_y f\left( {x,y} \right) = \mathop {\max }\limits\_y f\left( {{x^\*},y} \right)\]since we can view the inner max operation as being evaluated for a specific x value, and since \[{x^\*}\]is optimal, we can remove the min operation.

Now let use examine the situation where the min and max operations are switched such that \[\mathop {\max }\limits\_y \mathop {\min }\limits\_x f\left( {x,y} \right)\]. Focusing on the inner min operation: since the inner min operation is the minimum over all x values then \[\mathop {\min }\limits\_x f\left( {x,y} \right) \le f\left( {{x^\*},y} \right)\]. Since this inequality must hold for all values of y, the inequality is preserved when we take the max operation of y on both sides of the inequality. Therefore, we have that \[\mathop {\max }\limits\_y \mathop {\min }\limits\_x f\left( {x,y} \right) \le \mathop {\max }\limits\_y f\left( {{x^\*},y} \right)\]. Now substituting in the expression from above we have:\[\mathop {\max }\limits\_y \mathop {\min }\limits\_x f\left( {x,y} \right) \le \mathop {\min }\limits\_x \mathop {\max }\limits\_y f\left( {x,y} \right)\],

\*\***An open question remains regarding when the order of min and max can be exchanged**.\*\*

# Useful General Knowledge

This section contains general knowledge and tricks about different things that are useful

## Use Cholesky to Sample from Gaussian

There are a number of python functions that can be used to sample from a multivariate Gaussian. One technique is to sample from a standard Gaussian (i.e. zero mean, identity covariance), and then transform the sample such that it resembles being sampled from a Gaussian with mean m and covariance S. This can be achieved as follows (taken from “Gaussian Processes for Machine Learning slide 9”)

Figure\_\_500\_\_Figure 1. Cholesky Algorithm

Graphical user interface, text, application

Description automatically generated

This is the way it needs to be.