

Penalty Functions

- The Premise
- Quadratic Loss
- Problems and Solutions

The Premise

You may have noticed that the addition of constraints to an optimization problem has the effect of making it much more difficult.

The goal of **penalty functions** is to convert constrained problems into unconstrained problems by introducing an artificial penalty for violating the constraint.

The Premise

Consider this example: Suppose there is a freeway (like a toll freeway) that monitors when you enter and exit the road.

Instead of establishing a speed limit of 65 mph, they could use these rules:

- You can drive as fast as you want.
- If you drive under 65 mph you can use our road for free.
- Every mph you drive over 65 will cost you \$500.

The Premise

The previous example had no constraints – you can drive as fast as you want! But the effect of these rules would still be to limit drivers to 65mph. You can also control the likelihood of speeding by adjusting the fine.

Penalty functions work in exactly this way.

Initial Steps

We will be working with a very simple example:

$$\text{minimize } f(x) = 100/x$$

$$\text{subject to } x \leq 5$$

(With a little thought, you can tell that $f(x)$ will be minimized when $x = 5$, so we know what answer we should get!)

Before starting, convert any constraints into the form (expression) ≤ 0 . In this example, $x \leq 5$ becomes:

$$x - 5 \leq 0$$

Initial Steps

Once your constraints are converted, the next step is to start charging a penalty for violating them. Since we're trying to **minimize** $f(x)$, this means we need to **add value** when the constraint is violated.

If you are trying to maximize, the penalty will subtract value.

Quadratic Loss: Inequalities

With the constraint $x - 5 \leq 0$, we need a penalty that is:

- 0 when $x - 5 \leq 0$ (the constraint is satisfied)
- positive when $x - 5$ is > 0 (the constraint is violated)

This can be done using the operation

$$P(x) = \max(0, x - 5)$$

which returns the maximum of the two values, either 0 or whatever $(x - 5)$ is.

We can make the penalty more severe by using

$$P(x) = \max(0, x - 5)^2.$$

This is known as a **quadratic loss function**.

Quadratic Loss: Equalities

It is even easier to convert equality constraints into quadratic loss functions because we don't need to worry about the operation ($\max, g(x)$). We can convert $h(x) = c$ into $h(x) - c = 0$, then use

$$P(x) = (h(x) - c)^2$$

The lowest value of $P(x)$ will occur when $h(x) = c$, in which case the penalty $P(x) = 0$. This is exactly what we want.

Practice Problem 1

Convert the following constraints into quadratic loss functions:

a) $x \leq 12$

b) $x^2 \geq 200$

c) $2x + 7 \leq 16$

d) $e^{2x+1} \geq 9$

e) $4x^2 + 2^x = 12$

The Next Step

Once you have converted your constraints into penalty functions, the basic idea is to add all the penalty functions on to the original objective function and minimize from there:

$$\text{minimize } T(x) = f(x) + P(x)$$

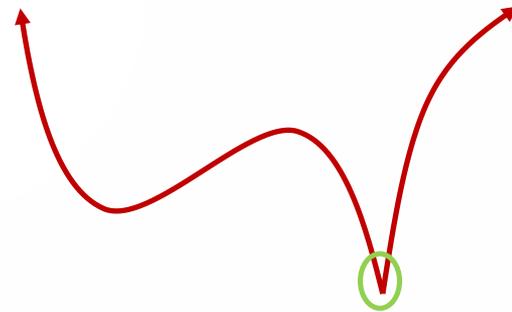
In our example,

$$\text{minimize } T(x) = 100/x + \max(0, x - 5)^2$$

A Problem

But... it isn't quite that easy.

The addition of penalty functions can create severe slope changes in the graph at the boundary, which interferes with typical minimization programs.



Fortunately, there are two simple changes that will alleviate this problem.

First Solution: r

The first is to multiply the quadratic loss function by a constant, r . This controls how severe the penalty is for violating the constraint.

The accepted method is to start with $r = 10$, which is a mild penalty. It will not form a very sharp point in the graph, but the minimum point found using $r = 10$ will not be a very accurate answer because the penalty is not severe enough.



First Solution: r

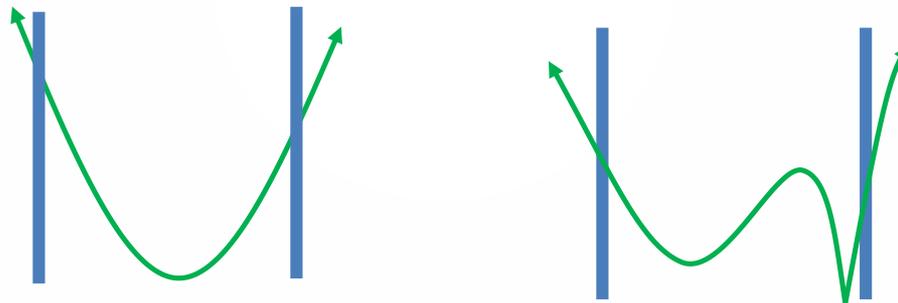
Then, r is increased to 100 and the function minimized again starting from the minimum point found when r was 10. The higher penalty increases accuracy, and as we narrow in on the solution, the sharpness of the graph is less of a problem.

We continue to increase r values until the solutions converge.



Second Solution: Methods

The second solution is to be thoughtful with how we minimize. The more useful minimization programs written in unit 2 were interval methods. The program started with an interval and narrowed it in from the endpoints.



With a severe nonlinearity, interval methods have a tendency to skip right over the minimum.

Second Solution: Methods

For this reason, interval methods are generally not ideal for penalty functions. It is better to use methods that take tiny steps from a starting point, similar to the “brute force” methods we used in 1-variable, or any of the methods we used in 2-variable minimization.

It is also important when using penalty functions to run the program a few times from various starting points to avoid local extremes.

Practice Problem 2

Write a program that, given an initial point and a function,

- 1) uses the derivative to determine the direction of the minimum
- 2) takes small steps in that direction until the function value increases
- 3) decreases the step size to narrow in on the minimum
- 4) reports the minimum value

Test, document and save your code!

The Final Step

Now, back to the example:

$$\begin{aligned} &\text{minimize } f(x) = 100/x \\ &\text{subject to } x \leq 5 \end{aligned}$$

has become

$$\text{minimize } T(x) = 100/x + r \cdot (\max(0, x - 5))^2$$

The initial point must be chosen **in violation of** the constraint, which is why these are known as “exterior penalty functions”. We’ll start with $x = 20$.

Practice Problem 3

Using your step minimization program, minimize

$$f(x) = 100/x + 10 \cdot (\max(0, x - 5))^2$$

from starting point 20. Call your answer “a”.

Then, minimize

$$f(x) = 100/x + 100 \cdot (\max(0, x - 5))^2$$

from starting point a.

Repeat for $r = 1000$ and $r = 10,000$.

Practice Problem 4

Write a function that will carry out successive iterations of raising the value of r and closing the interval boundaries. Check your loop with the previous problem, then use it to solve this problem:

$$\text{minimize } f(x) = 0.8x^2 - 2^x$$

$$\text{subject to } x \leq 4$$

Test different starting points to see the effect.

Practice Problem 5

Next, solve this problem:

$$\text{Minimize } f(x) = (x - 8)^2$$

$$\text{subject to } x \geq 10$$

(Be careful with that \geq ! it will affect both your $g(x)$ and your starting point.)

A Note About Exterior Penalty Functions

Because exterior penalty functions start outside the feasible region and approach it from the outside, they only find extremes that occur on the boundaries of the feasible region. They will not find interior extremes.

In order to accomplish that, these are often used in combination with interior penalty functions... next lesson!