Evolution as Stochastic Gradient Descent

Evolution is a gradient descent algorithm[1]. An organism can be thought of as a set of many, many parameters, encoded first in genes, then later in the organism's structure and form. The objective is simple: to persist. The external environment provides the valleys and peaks of the cost function, which is a function of the organism's parameters. It measures how good this organism is at surviving. Too costly and you will die. Less costly and you will prosper. But the parameters that compose an organism are not static. They are constantly and slowly tweaked by random mutations.

Most of these changes will inevitably be in a bad direction. This is because the organism has already had time to adapt; it quickly reaches niches at which point many of the possible mutations will harm it. The environment is harsh to organisms that are expensive; higher cost means it will quickly die off and be culled from the genetic pool. But every once in a while, it gets a push in a good direction; lower in the cost-function landscape. These organisms tend to survive, and their less-cost-efficient ancestors will die. This is how a species adapts.

But the environment is not static either. It changes, whether due to the influence of abiotic causes or the environmental pressures exerted its fellow organisms. Remember that an organism does not exist in a vacuum; the environment is also constantly changed by the organisms evolving around it. In fact, if two organisms are too close, one of them will often die out due to the increased environmental pressures. These cause the cost-landscape to shift; organisms in a safe local minimum will have to change in order to stay in their local minimum. And for some unlucky critters, their local minimum will be pushed so high that they die out.

This process is inefficient: both slow and also *blind*. First, slow; it takes orders of thousands to hundreds of thousands of years to see noticeable differences (depending on how long a generation takes, and how big of a difference you want). This means that the gradient descent algorithm of an organism has a very low *step size*, and these steps take a long time to do. Many an extinction event has been caused by the cost-landscape shifting so quickly as to push organisms up across the "survival threshold" with nary a blink of an eye, evolutionarily speaking, to adapt. You might think of global warming, perhaps, or a stray meteor.

Second, blind. A gene can only "see" the gradient exactly where it lies on the cost-landscape. It is completely nearsighted. This is true for must gradient descent algorithms. But perhaps "see" is an overstatement; it "knows" (in only the most generous of ways) the gradient only by virtue of generations of random genetically drifting organisms, like little "probes", where weaklings perish and only the strong remain, in the direction of downwards slope.

Notice how difficult it is for costly organisms to survive! They are constantly outpaced by their neighbors, which might also exert negative environmental forces on them (because they take up more space, eat the available food, etc), and sooner or later the environment will wipe them out. So an organism really has no way to do audacious, multi-step "experiments." Every single step on the way to a large adaptation (like wings or eyes or a backbone) must, at the very least, *not harm* the organism's evolutionary fitness. This means an organism's path is limited to paths straight downwards in the landscape; it cannot go over ridges in search of a better (lower) life for its offspring, cannot cross vast swaths of peaks in order to find a much better minimum. This is why cheetahs cannot evolve guns on their elbows (despite this weapon being very good at killing prey, though take this with a grain of salt - I have yet to ask a cheetah what they think). To do so requires a series of large evolutionary changes that will no doubt be bad at the outset for the cheetah. Vaguely gun-shaped elbows, a requisite along the evolutionary path to true gun-elbows, do not advance its current goal, which is to run and catch prey in its jaws.

So where is our algorithm? It is a gradient descender with a fixed (tiny) step size that requires thousands of years to run, is highly stochastic locally (it travels in random directions, only the better ones survive), and is unforgiving in the sense that it is highly rigid in allowed paths. It is, in other words, not very good. What we should expect is that it has *missed* many, many, better minima. The notion that evolution always converges on the best strategy is a myth. This gives us hope. It means that there are much more efficient, much better ways to make organisms.

Yet it has had billions of years to run, and has access to an unbelievably vast number of parameters (parameters here, of course, being an analogy for morphological characteristics, each of which depends on a vast array of protein structures, each of which can mutate, even making new structures and new parameters), and by virtue of its stochasticity (albeit highly local), has birthed a truly astounding array of life-forms. The shifting environment is largely to blame for the diversity of life; a constantly changing landscape allows individual organisms to explore much wider in the state-space of possible forms.

But evolution has endowed us with a power much more potent than itself. We have the power to plan, to search possible paths *before we do them*. We do this by imagining things, by invoking much more potent search algorithms than just blind gradient descent. We see patterns; we can predict what things will be good before we implement them. There is, of course, an unmistakeable air of trial-and-error to the whole thing - like evolution, we fail much more than we succeed. But we do not make our innovations at random - our search is not blind, per se, but informed by our past experiences and the patterns we see in the world. We can plan for multiple steps before realizing the result, and as such are not limited by the hills and valleys of the cost function. We can explore, more-or-less informedly, to find minima unreachable by blind gradient descent. This is why we invented the gun, the aeroplane, and the internet. This algorithm is what we call technology.

---

1. The astute among you might remark that evolution is actually an evolutionary algorithm (no kidding!). In this sense, it is not a gradient descent algorithm, since it never actually calculates a gradient. I must say, you make a compelling point. I'm twisting the idea of gradient descent a little here, though I'd argue that evolutionary algorithms are actually a very stochastic subset of gradient descent - one that approximates a downward gradient through random sampling. ↩