9/15 Notes

Friday, September 15, 2023 8:51 AM

This chapter is about finding the solution to

Note if we want max ← $\min(f(x)) \Leftrightarrow \max(-f(x))$

$$\min\{f(\mathbf{x}) \mid \mathbf{x} \in \mathbb{R}^n\}$$

Reasons why this is hard:

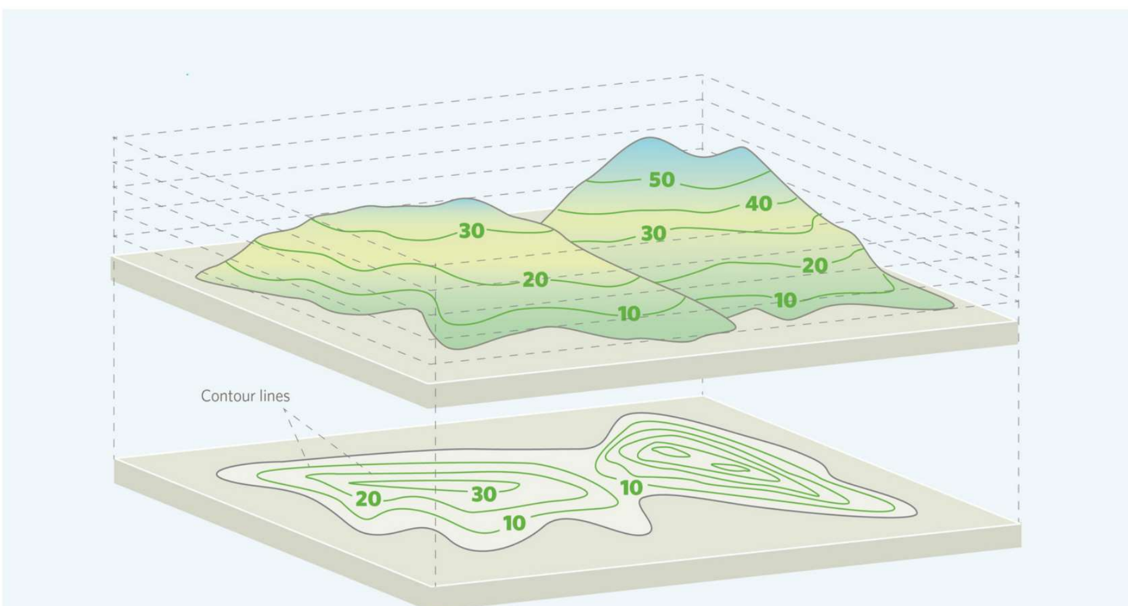
- It might be a very difficult task to solve the set of (usually nonlinear) equations $\nabla f(\mathbf{x}) = \mathbf{0}$.
- There might be an infinite number of stationary points and the finding the one corresponding to the minimal function value is an optimization problem which by itself might be as difficult as the original problem.

→ we had a similar problem on the HW but it was set up in a convient way to solve

Descent

Say we want to find the minimum w/o taking $\nabla f(x) = 0$

The Foggy Mountain



- We want to decend the Mountain as fast as possible.

→ At each step we want to go in the direction of greatest slope

→ We can't see because of the fog, so we only know at the point we're at

→ At each step we want to go in the direction of greatest slope

→ We can't see because of the fog, so we only know at the point we're at

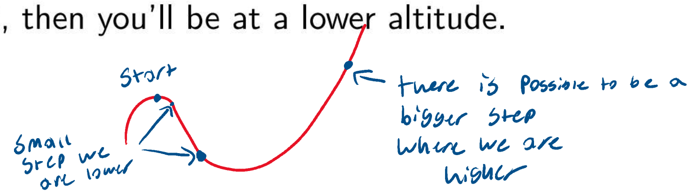
Descent direction

Let $f: \mathbb{R}^n \to \mathbb{R}$ be a continuously differentiable function over $\mathbb{R}^n$. A nonzero vector $\mathbf{d} \in \mathbb{R}^n$ is a *descent direction* of f at $\mathbf{x}$ if the *directional derivative* $f'(\mathbf{x}; \mathbf{d}) = \nabla f(\mathbf{x})^\top \mathbf{d}$ is negative.

→ We want to go in the direction where the ground is sloping down

Lemma: Let f be a continuously differentiable function over an open set U, and let $\mathbf{x} \in U$. Suppose that $\mathbf{d}$ is a descent direction of f at $\mathbf{x}$. Then there exists $\varepsilon > 0$ such that $f(\mathbf{x} + t\mathbf{d}) < f(\mathbf{x})$ for any $t \in (0, \varepsilon]$.

- This means that there's a t such that if you start at $\mathbf{x}$ and walk in direction $\mathbf{d}$ for a distance t, then you'll be at a lower altitude.



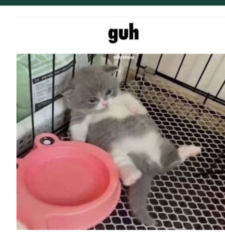
Our algorithm to do our descent

Initialization: Pick $\mathbf{x}_0 \in \mathbb{R}^n$ arbitrarily. General step: For any $k = 0, 1, 2, \ldots$ do:

1. Pick a descent direction $\mathbf{d}_k$.
2. Find a stepsize t_k satisfying $f(\mathbf{x}_k + t_k\mathbf{d}_k) < f(\mathbf{x}_k)$.
3. Set $\mathbf{x}_{k+1} = \mathbf{x}_k + t_k\mathbf{d}_k$.
4. If a stopping criterion is satisfied, then STOP and $\mathbf{x}_{k+1}$ is the output.

Explaining what this means:

Stepsize



There are several ways to find the stepsize:

- Let the stepsize be constant.
  - How do we find the constant? ← GUH :(
- Exact line search: let t_k be a minimizer of f along the ray $\mathbf{x}_k + t_k\mathbf{d}_k$.
  - Not always possible to find the exact minimizer. ← GUH :(
- Backtracking: pick three parameters: an initial guess $s > 0$, and $\alpha \in (0, 1), \beta \in (0, 1)$. Then the stepsize is $t_k = s\beta^{i_k}$, where i_k is the smallest non-negative integer such that $f(\mathbf{x}_k) - f(\mathbf{x}_k + s\beta^{i_k}\mathbf{d}_k) \geq -\alpha s\beta^{i_k}\nabla f(\mathbf{x}_k)^\top\mathbf{d}_k$.
  - Compromise that finds a "good enough" stepsize. ← YAY ^^
  - A theorem guarantees the existence of i_k. ← in book

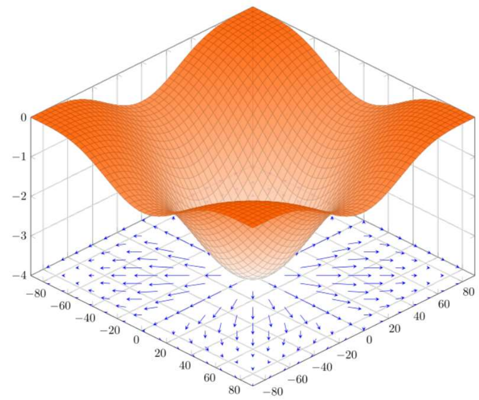
→ This would take forever by hand

→ This would take forever by hand

Picking a direction $\mathbf{d}$

In the gradient method the descent direction is $\mathbf{d}_k = -\nabla f(\mathbf{x}_k)$.

- The direction of the gradient at $\mathbf{x}$ is the direction in which $f(\mathbf{x})$ increases the most from the point $\mathbf{x}$. ← So we take negative to see decreasing



The Gradient Method

Input: tolerance parameter $\varepsilon > 0$. Used since finding exact solution is difficult. We want to specify when the change is close enough to 0 to be happy :)

Initialization: Pick $x_0 \in \mathbb{R}^n$ arbitrarily. General step: For any $k = 0, 1, 2, \ldots$ do:

1. Pick a stepsize t_k by a line search procedure on the function $g(t) = f(\mathbf{x}_k - t\nabla f(\mathbf{x}_k))$.
2. Set $\mathbf{x}_{k+1} = \mathbf{x}_k - t_k \nabla f(\mathbf{x}_k)$.
3. If $\|\nabla f(\cancel{\mathbf{x}_{k+1}})\| \le \varepsilon$, then STOP and $\mathbf{x}_{k+1}$ is the output.
   x_{k+1}

Example

Find $\min_{\mathbf{x},\mathbf{y}} (x-2)^2 + (y-1)^2$. ← Look at it we know $(2,1)$

Initialization: Pick $\mathbf{x}_0 \in \mathbb{R}^n$ arbitrarily. General step: For any $k = 0, 1, 2, \ldots$ do:

1. Pick a descent direction $\mathbf{d}_k$.
2. Find a stepsize t_k satisfying $f(\mathbf{x}_k + t_k\mathbf{d}_k) < f(\mathbf{x}_k)$.
3. Set $\mathbf{x}_{k+1} = \mathbf{x}_k + t_k\mathbf{d}_k$.
4. If a stopping criterion is satisfied, then STOP and $\mathbf{x}_{k+1}$ is the output.

1) Initial guess: $(0,0)$

2) Step size: minimize $\{f(\vec{x}_0 - t\nabla f(\vec{x}_0))\}$

$\Rightarrow \nabla f(x,y) = \begin{pmatrix} 2(x-2) \\ 2(y-1) \end{pmatrix}$ $\nabla f(0,0) = \begin{pmatrix} -4 \\ -2 \end{pmatrix}$

$\Rightarrow f([(0,0) - t(-4,-2)]) = f(4t, 2t) = (4t-2)^2 + (2t-1)^2$

$\Rightarrow$ minimize $\{(4t-2)^2 + (2t-1)^2\} \Rightarrow t = 1/2$

usually using derivative $\frac{d}{dt}f(t)=0$

3) Find next iter: $\vec{x}_1 = \vec{x}_0 - \frac{1}{2}(-4,-2) = (0,0) - (-2,-1) = (2,1)$

Note we should have set a value for ε at the start

4) Check stopping criterion: $\|\nabla f(2,1)\| = \left\| \begin{bmatrix} 2(2)-4 \\ 2(1)-2 \end{bmatrix} \right\| = \left\| \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right\| = 0 < \varepsilon$ for $\varepsilon > 0$