Exercises in Convex Optimization
Lecture 9

Michel Baes, Patrick Cheridito

November 21, 2018

The Golden Search We want to minimize the convex function $f : [a, b] \to \mathbb{R}$. We have at our disposal an oracle of order 0: given an input point in $[a, b]$, this oracle only returns the value of the f at that point, but none of its subgradients. We assume that f has only one minimizer x^* on $[a, b]$. We compare in this exercise two minimization algorithms: the trisection method, which intuitively seems the most natural, and the golden section method, which performs considerably better.

1. Consider Algorithm 1.

Algorithm 1 TRISECTIONMETHOD$(f, [a, b])$

$a_0 := a$, $b_0 := b$, $\Delta_0 := [a_0, b_0]$.
for $k \geq 0$
- Compute $f_k^- := f\left(\frac{2a_k}{3} + \frac{b_k}{3}\right)$ and $f_k^+ := f\left(\frac{a_k}{3} + \frac{2b_k}{3}\right)$.
- **if** $f_k^- \leq f_k^+$, set $a_{k+1} := a_k$ and $b_{k+1} := \frac{a_k}{3} + \frac{2b_k}{3}$.
  else set $a_{k+1} := \frac{2a_k}{3} + \frac{b_k}{3}$ and $b_{k+1} := b_k$.
  end if
- Set $\Delta_{k+1} := [a_{k+1}, b_{k+1}]$.

end for

   Show that this algorithm converges linearly with convergence ratio $\sqrt{2/3}$: prove that after N function evaluations, we have $|x_N - x^*| \leq \left(\frac{2}{3}\right)^{\lfloor N/2 \rfloor} \cdot |b - a|$, where x_N is an arbitrary point in $\Delta_{\lfloor N/2 \rfloor}$. Why is the above method obviously not the optimal one?

2. Let $\lambda := \frac{\sqrt{5}-1}{2}$ be the *golden number*, for which we have $\lambda + 1 = \frac{1}{\lambda}$, and consider Algorithm 2.

   Show that, except at iteration $k = 0$, one of the two values $f_k^- := f(\lambda a_k + (1-\lambda)b_k)$ and $f_k^+ := f((1-\lambda)a_k + \lambda b_k)$ has already been computed (i.e. coincides with some f_j^+ or f_j^- for a $j < k$).

   Show that this algorithm converges linearly with convergence ratio λ: prove that after $N \geq 2$ function evaluations, we have $|x_N - x^*| \leq \lambda^{N-1}|b - a|$, where x_N is an arbitrary point in Δ_{N-1}.

`Lecture_10/GoldenSection`

Algorithm 2 GoldenSectionMethod($f, [a, b]$)

$a_0 := a$, $b_0 := b$, $\Delta_0 := [a_0, b_0]$, $\lambda := \frac{\sqrt{5}-1}{2}$.
for $k \geq 0$
- Compute $f_k^- := f(\lambda a_k + (1-\lambda)b_k)$ and $f_k^+ := f((1-\lambda)a_k + \lambda b_k)$.
- **if** $f_k^- \leq f_k^+$, set $a_{k+1} := a_k$ and $b_{k+1} := (1-\lambda)a_k + \lambda b_k$.
- **else** set $a_{k+1} := \lambda a_k + (1-\lambda)b_k$ and $b_{k+1} := b_k$.
- **end if**
- Set $\Delta_{k+1} := [a_{k+1}, b_{k+1}]$.

end for

Do "super-smooth" functions exist? Let $\|\cdot\|$ be any norm on $\mathbb{R}^n$. Let $f\colon D \to \mathbb{R}$ be a twice continuously differentiable function that is convex on the convex set D with $D \subseteq \mathbb{R}^n$ and $\mathrm{int}(D) \neq \emptyset$. Suppose that there exist $\alpha > 0$ and $k > 2$ such that

$$0 \leq f(x+h) - f(x) - \langle f'(x), h\rangle \leq \alpha\|h\|^k. \tag{1}$$

Prove that f is affine, that is, that there exist $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$ such that $f(x) = \langle a, x\rangle + b$.

(Hint: Use mean value theorem.)

When equation (1) holds with $k = 2$, we call f a *smooth function*. When we can take $k = 1$, we call f *Lipschitz continuous*. We call functions that satisfy (1) for a number $k \in [1, 2]$ *Hölder-smooth*. This exercise shows that the inequality is rather uninteresting for $k > 2$.

`Lecture_10/HolderSmooth`

Convergence rate of steepest descent method Let the twice continuously differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ be strongly convex, i.e., there exists $\mu > 0$ such that

$$f(y) \geq f(x) + \langle f'(x), y - x\rangle + \frac{\mu}{2}||y-x||_2^2$$

for all $x, y \in \mathbb{R}^n$. Let $x_0 \in \mathbb{R}^n$ and denote $S = \{x \mid f(x) < f(x_0)\}$. Suppose that S is not empty.

1. Show that S is bounded and therefore that there exists $L > 0$ such that

$$f(y) \leq f(x) + \langle f'(x), y - x\rangle + \frac{L}{2}||y-x||_2^2$$

for all $x, y \in S$.

2. Let $x \in S$ and denote $t^+ = \mathrm{argmin}\{f(x - tf'(x)) \mid t > 0\}$ and $x^+ = x - t^+ f'(x)$. Show that following holds:

$$f(x_+) - f(x^*) \leq \left(1 - \frac{\mu}{L}\right)[f(x) - f(x^*)],$$

where $x^* = \mathrm{argmin}\{f(x) \mid x \in S\}$.
(Hint: Start by justifying why $x_+ \in S$. You can use in an intermediate step that $f(x^*) \geq \min_t f(x) - t||f'(x)||_2 + \frac{\mu}{2}t^2$, but you need to justify why this inequality is true, to compute the right-hand side, and to figure out what you can do with it.)

`Lecture_10/SteepestDescentConvergenceRate`