**Large-scale convex optimisation**

Sebastian Banert **Exercise 6** 29 December 2020

---

**Problem 1.** Let $\mathcal{H}$ be a finite-dimensional inner-product space, let $f\colon \mathcal{H} \to \overline{\mathbb{R}}$ be convex and proper, and let $C \subseteq \operatorname{int}(\operatorname{dom} f)$ be non-empty, closed, and convex. We want to minimise $f(x)$ over $x \in C$. Given $x_0 \in C$, the *projected subgradient method* is given by the following iteration for $n \geq 0$:

- choose $s_n \in \partial f(x_n)$,
- choose $\gamma_n > 0$,
- set $x_{n+1} := \operatorname{Proj}_C(x_n - \gamma_n s_n)$, where $\operatorname{Proj}_C$ denotes the projection on the set $C$.

Show that, for all $x \in C$,

$$\|x_{n+1} - x\|^2 \leq \|x_n - x\|^2 - 2\gamma_n(f(x_n) - f(x)) - \|x_{n+1} - (x_n - \gamma_n s_n)\|^2 + \gamma_n^2\|s_n\|^2. \quad (1)$$

Which parts of the proof of the subgradient method have to be adapted to show the convergence of the projected subgradient method?

**Problem 2.** Assume that you know that there exists $m \in \mathbb{R}$ such that $f(x) \geq m$ for all $x \in \mathcal{H}$. If you set

$$\gamma_n := \frac{f(x_n) - m}{\|s_n\|^2}$$

for all $n \geq 0$ in the subgradient method (or the projected subgradient method), show that

$$\sum_{n=0}^{N-1} \frac{(f(x_n) - f(\bar{x}))^2 - (f(\bar{x}) - m)^2}{\|s_n\|^2} \leq \|x_0 - \bar{x}\|^2$$

for all $N \geq 0$, where $\bar{x}$ is a minimiser of $f$.

Note: If, for some reason, you know the optimal value $f(\bar{x})$ of your optimisation problem and want to find a solution $\bar{x}$, you can set $m = f(\bar{x})$. This minimises the expression $-2\gamma_n(f(x_n) - f(x)) + \gamma_n^2\|s_n\|^2$ in eq. (1). If you don't know $f(\bar{x})$, then the rule here does not guarantee $f(x_n) \to f(\bar{x})$, but it can be a good rule of thumb as long as $f(x_n) - f(\bar{x})$ is big compared with $f(\bar{x}) - m$. You can read more about this and the subgradient method in Polyak's book [1, Chapter 5.3].

**Problem 3.** Implement the projected subgradient algorithm in a programming language of your choice (e.g., Julia, Python, GNU Octave, …). Make sure that I only need free (open source) software to execute your code. (Matlab is not open source; if you want to use Matlab syntax, make sure that your code runs with GNU Octave.)

Use your implementation to numerically solve the shortest-path problem from the lecture:

$$\begin{aligned} \min \quad & \sum_{k=1}^{N} \|x_k - x_{k-1}\| && \text{over } x_0, \ldots, x_N \in \mathbb{R}^n \\ \text{s.t.} \quad & x_0 = a, \qquad x_N = b. \end{aligned}$$

with $n = 2$ and $a = (0,0) \in \mathbb{R}^2$ and $b = (1,2) \in \mathbb{R}^2$.

- Choose the initial path randomly.
- Don't be confused by the overlapping notation in the problem formulation and the subgradient algorithm.
- Use $\gamma_n = \frac{a}{(b+n)\|s_n\|}$ for a reasonable (manually tuned) choice of $a, b > 0$ and the $\gamma_n$ from problem 3 with $m = f(\bar{x}) = \|a - b\|$ and $m = 0$.
- Plot the function values $f(x_n)$, $n = 0, 1, \ldots$ and the weighted averages from the ergodic convergence rate for each choice of stepsizes.
- Check that eq. (1) is satisfied for each step (using for $x$ the point with the lowest objective function value you found).

Visualise the shortest path found by your algorithm. Change the norm in the objective function to $\|\cdot\|_1$ and $\|\cdot\|_\infty$, given by

$$\|(a,b)\|_1 = |a| + |b|, \qquad \|(a,b)\|_\infty = \max\{|a|, |b|\},$$

and visualise the shortest paths for those. Send me your code and all the calculations you need for your implementation.

## References

[1] Boris T. Polyak, *Introduction to optimization*, Optimization Software, 1987.