9 Projection operator, and projected (sub)gradient methods

Projection operator If $C \subset \mathbb{R}^n$ is a closed convex set, the Euclidean projection on C is defined by

$$\mathbf{proj}_C(y) = \operatorname*{argmin}_{x \in C} \|x - y\|_2^2. \tag{1}$$

Observe that the projection mapping satisfies

$$\langle y - \mathbf{proj}_C(y), x - \mathbf{proj}_C(y) \rangle \le 0 \ \ \forall x \in C. \tag{2}$$

(This is precisely the optimality condition written for (1)). The inequality above can in fact be summarized as $y - \mathbf{proj}_C(y) \in N_C(\mathbf{proj}_C(y))$. It immediately follows from (2) that $\mathbf{proj}_C$ satisfies

$$\| \mathbf{proj}_C(y) - \mathbf{proj}_C(z)\|_2^2 \le \langle y - z, \mathbf{proj}_C(y) - \mathbf{proj}_C(z) \rangle$$

which implies, that $\mathbf{proj}_C$ is nonexpansive

$$\| \mathbf{proj}_C(y) - \mathbf{proj}_C(z)\|_2 \le \|y - z\|_2$$

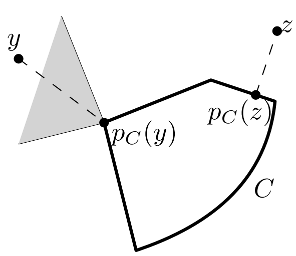

Figure 1: Projection operator (denoted p_C in the figure) on a closed convex set C. Grey shaded region is the normal cone $N_C(\mathbf{proj}_C(y))$.

Projected (sub)gradient method Consider the constrained minimization problem

$$\min_{x \in C} f(x).$$

The projected (sub)gradient method has iterates

$$x_{k+1} = \mathbf{proj}_C(x_k - t_k g_k) \tag{3}$$

where $g_k \in \partial f(x_k)$; if f is smooth then of course we have $g_k = \nabla f(x_k)$. Note that the fixed point equation of the iterates (3) is $x^* = \mathbf{proj}_C(x^* - tg(x^*))$ where $g(x^*) \in \partial f(x^*)$, which is equivalent, by the properties of the projection operator, that $-g(x^*) \in N_C(x^*)$, and so in particular $0 \in \partial(f + I_C)(x^*)$.

One can easily modify the convergence proofs performed in the unconstrained case, to obtain quantitative rates on the convergence, depending on the properties of f. The rates we obtain are exactly the same as in the unconstrained case. We briefly summarize the changes needed for the convergence proofs:

- Nonsmooth f (subgradient method): we use the nonexpansive property of $\mathbf{proj}_C$ to get

$$\|x_{k+1} - x^*\|_2^2 = \| \mathbf{proj}_C(x_k - t_k g_k) - \mathbf{proj}_C(x^*)\|_2^2 \leq \|x_k - t_k g_k - x^*\|_2^2$$

  and the rest of the proof is the same as the standard subgradient method.

- f is L-smooth (gradient method): Write $\tilde{x} = x - t\nabla f(x)$ and $x^+ = \mathbf{proj}_C(\tilde{x})$. We have

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

  By expressing $\nabla f(x) = -(\tilde{x} - x)/t = -(\tilde{x} - x^+ + x^+ - x)/t$ and using the property (2) about the projection we get

$$f(x^+) \leq f(x) - \frac{1}{t}\|x^+ - x\|_2^2(1 - Lt/2) = f(x) - \frac{1}{2t}\|x^+ - x\|_2^2.$$

  The rest of the proof is exactly the same.

- f is m-strongly and L-smooth: using the nonexpansive property of $\mathbf{proj}_C$, and the fact that $x^* = \mathbf{proj}_C(x^* - t\nabla f(x^*))$, we have $\|x^+ - x^*\|_2 \leq \|x - x^* - t(\nabla f(x) - \nabla f(x^*))\|_2$, and the proof follows exactly the same lines as in Lecture 3.

The projected gradient method is only a suitable method when the projection map $\mathbf{proj}_C$ can be easily computed, e.g., when $C = \{x : \|x\|_\infty \leq r\}$, $C = \{x : x \geq 0 \text{ and } \sum_{i=1}^n x_i = 1\}$, etc. Computing the projection on a general convex set however is itself a nontrivial convex optimization problem.

EXERCISE: Give explicit expressions for the projection maps on the following sets: $\mathbb{R}^n_+ = \{x \in \mathbb{R}^n : x \geq 0\}$, $\{x \in \mathbb{R}^n : \|x\|_\infty \leq 1\}$, $\{x \in \mathbb{R}^n : \|x\|_2 \leq 1\}$.

EXERCISE: Explain how to efficiently compute the projection on the unit simplex $C = \{x \in \mathbb{R}^n : x \geq 0 \text{ and } \sum_{i=1}^n x_i = 1\}$. (Hint: consider solving the dual problem).