LINMA2471 – Optimization models ans methods II
Notes from the 6th lecture

Antoine Aspeel Pierre-Paul Mouchet Guillaume Olikier

October 21, 2015

1 Reminder on solvers

This section brings complements to what was discussed in the first lecture about solvers.

Solvers	Problems	Integer variables
CPLEX,GUROBI	linear optimization and convex quadratic optimization	yes
KNITRO, SNOPT, MINOS	nonlinear optimization	yes for KNITRO but loss of efficiency
BARON	global optimization	no

2 Gradient method for unconstrained problems

Let us recall basic definitions from last lecture.

Definition 1. Given $L > 0$, we say $f : D \subseteq \mathbb{R}^n \to \mathbb{R}$ has *L-Lipschitz gradient* if and only if $f \in C^1(D)$ and

$$\|\nabla f(x) - \nabla f(y)\| \leq L \|x - y\|$$

for all $x, y \in D$. We denote $C_L^{1,1}(D)$ the set of such functions. We also define

$$F_L^{1,1}(D) = \{f \in C_L^{1,1}(D) \mid f \text{ is convex}\}.$$

Given $f \in C^1$, we denote $T_y^1(x) = f(y) + \nabla f(y)^T(x - y)$ the first-order Taylor expansion of f around y evaluated at point x.

2.1 Gradient method for functions of $C_L^{1,1}$

Last week, we studied the gradient method for the general problem

$$\min_{x \in \mathbb{R}^n} f(x)$$

where $f \in C_L^{1,1}(\mathbb{R}^n)$. We proved that $\frac{1}{L}$ was actually the best step length choice and that it guaranteed

$$\min_{0 \leq i \leq N} \|\nabla f(x_i)\| \leq \sqrt{\frac{2L(f(x_0) - f(x^*))}{N + 1}}. \tag{1}$$

Observe that this inequality

- is scaling independent,
- doesn't say anything about the values of f.

One can show that this inequality is not improvable.

Let us recall the way we obtained the above inequality.

Lemma 1 (Quadratic bounds in $C_L^{1,1}$)**.** *The following conditions are equivalent :*

(a) $f \in C_L^{1,1}(D)$,

(b) $f \in C^1(D)$ *and* $|f(y) - T_x^1(y)| \leq \frac{L}{2}\|x - y\|^2$ *for all* $x, y \in D$.

Proof. See the fourth exercises session. □

From this lemma, we concluded that $\frac{1}{L}$ is the optimal step length.

Theorem 2 (Decrease guarantee)**.** *Let* $f \in C_L^{1,1}$. *Denote* $x^+ = x - \frac{1}{L}\nabla f(x)$ *the next iterate. Then*

$$f(x) - f(x^+) \geq \frac{\|\nabla f(x)\|^2}{2L}.$$

Proof. Use the upper bound of lemma 1 with $y = x^+$. □

Actually there exist a family of functions which can be as closed of this bound as you want.

Finally theorem 2 leads to the inequality (1).

2.2 Gradient method for functions of $F_L^{1,1}$

Let us now consider the same problem with the additional assumption $f \in F_L^{1,1}(\mathbb{R}^n)$. Lemma 1 can be improved as follows.

Lemma 3 (Quadratic bounds in $F_L^{1,1}$)**.** *The following conditions are equivalent :*

(a) $f \in F_L^{1,1}(D)$,

(b) $f \in C^1(D)$ *and* $T_y^1(x) \leq f(x) \leq T_y^1(x) + \frac{L}{2}\|x - y\|^2$ *for all* $x, y \in D$.

Proof. See the fourth exercises session. □

Lemma 4. *Let* $f \in C_L^{1,1}$. *For any optimal solution* x^* *and any* x,

$$\frac{\|\nabla f(x)\|^2}{2L} \leq f(x) - f(x^*) \leq \frac{L}{2}\|x - x^*\|^2.$$

Proof. The first inequality follows from theorem 2 since $f(x^+) \geq f(x^*)$. The second inequality follows from lemma 1 applied with $x = x^*$. Indeed since $f \in C^1$ and x^* is a local extremum, we have $\nabla f(x^*) = 0$. □

Theorem 5 (Convergence of $\frac{1}{L}$-gradient method for $F_L^{1,1}$)**.** *Let* $f \in F_L^{1,1}$. *For any iterate* x_N *and any* x,

$$f(x_N) - f(x^*) \leq \frac{L}{2N}\|x_0 - x^*\|^2.$$

Proof. We start from theorem 2:

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

Since $f \in C^1$ and f is convex, we have

$$f(x^*) \geq f(x) + \nabla f(x)^T(x^* - x),$$

right-hand side being tangent equation around x. Combining these two inequalities yields

$$f(x^+) \leq f(x^*) + \nabla f(x)^T(x - x^*) - \frac{1}{2L}\|\nabla f(x)\|^2.$$

Now, observe that[1]

$$\nabla f(x)^T(x-x^*) - \frac{1}{2L}\|\nabla f(x)\|^2 = \frac{L}{2}\left(\|x-x^*\|^2 - \left\|x-x^*-\frac{1}{L}\nabla f(x)\right\|^2\right).$$

Noting that $x - \frac{1}{L}\nabla f(x) = x^+$, we obtain

$$f(x^+) - f(x^*) \leq \frac{L}{2}\left(\|x-x^*\|^2 - \left\|x^+-x^*\right\|^2\right).$$

So given $N \in \mathbb{N}$, we have for all $i \in \{0, ..., N-1\}$

$$f(x_{i+1}) - f(x^*) \leq \frac{L}{2}\left(\|x_i-x^*\|^2 - \|x_{i+1}-x^*\|^2\right).$$

Summing those N inequalities yields

$$\begin{aligned}
\sum_{i=0}^{N-1} f(x_{i+1}) - Nf(x^*) &\leq \frac{L}{2}\sum_{i=0}^{N-1}\left(\|x_i-x^*\|^2 - \|x_{i+1}-x^*\|^2\right) \\
&= \frac{L}{2}\left(\|x_0-x^*\|^2 - \|x_N-x^*\|^2\right) \\
&\leq \frac{L}{2}\|x_0-x^*\|^2.
\end{aligned}$$

Notice now that $f(x_N) \leq f(x_i)$ for all $i \in \{0, ..., N-1\}$ so that

$$Nf(x_N) \leq \sum_{i=1}^{N} f(x_i).$$

Using this in the last inequality, we finally get

$$f(x_N) - f(x^*) \leq \frac{L}{2N}\|x_0-x^*\|^2. \qquad \square$$

Among all methods with

$$x_k \in \operatorname{span}\{x_0, \nabla f(x_0), ..., \nabla f(x_{N-1})\},$$

none of them can guarantee better than

$$f(x_N) - f(x^*) \leq \frac{3}{32}L\frac{\|x_0-x^*\|^2}{(N+1)^2}$$

for dimension greater or equal to $2N+1$.

3 Gradient method for constrained problems

We now add constraints. We consider problems of the following form:

$$\min_{x\in C} f(x) \qquad \text{with} \quad f \in C_L^{1,1}(C)$$

and

$$\min_{x\in C} f(x) \qquad \text{with} \quad f \in F_L^{1,1}(C).$$

We assume $C \subseteq \mathbb{R}^n$ is a convex and closed set. This implies that the orthogonal projection on C

$$P_C : \mathbb{R}^n \to C : x \mapsto P_C(x)$$

is well defined and unique.

[1] Apply $\|a-b\|^2 = \|a\|^2 - 2a^Tb + \|b\|^2$ to $a = x - x^*$ and $b = \frac{1}{L}\nabla f(x)$. Ok, it's a trick.

Definition 2. Given $f \in C_L^{1,1}(C)$, we say that x^* is a *stationary point* of the problem $\min_{x \in C} f(x)$ if and only if

$$\nabla f(x^*)^T (x - x^*) \geq 0$$

for all $x \in C$.

This definition can be intuitively interpreted as follows: adding $f(x^*)$ on both sides brings up the first-order Taylor expansion of f around x^*, which is closed to $f(x)$. So this definition essentially means $f(x) \geq f(x^*)$.

Note that if $x^* \in \operatorname{int} C$, then necessarily $\nabla f(x^*) = 0$. Indeed, if $\nabla f(x^*) \neq 0$ and $x^* \in \operatorname{int} C$, we can choose x such that $\nabla f(x^*)$ and $x - x^*$ are of opposite directions. Consequently, their scalar product is negative which contradicts the definition of x^*. This implies that $\nabla f(x^*) = 0$.

Theorem 6. *Under the above assumptions, if x^* is a local minimum, then x^* is stationary.*

Theorem 7. *When f is convex, stationarity implies optimality.*

Let us now present the gradient method for constrained problems. The principle is the following:

1. at each step, minimize the quadratic upper bound on set C,
2. which is equivalent to projecting the true minimum of the quadratic upper bound on set C.

Let us show this equivalence. Statements mean

1. choose x^+ minimizing $f(x) + \nabla f(x)^T (x^+ - x) + \frac{L}{2} \|x - x^+\|^2$ over C, where we can ignore the constant term $f(x)$ in the minimization problem,
2. choose x^+ minimizing $\left\|x^+ - (x - \frac{1}{L} \nabla f(x))\right\|^2$ over C. Notice that we can develop

   $$\left\|x^+ - (x - \frac{1}{L}\nabla f(x))\right\|^2 = \left\|(x^+ - x) + \frac{1}{L}\nabla f(x)\right\|^2 = \left\|x^+ - x\right\|^2 + \frac{2}{L}(x^+ - x)^T \nabla f(x) + \frac{1}{L^2}\left\|\nabla f(x)\right\|^2,$$

   which is equivalent to 1 since we can ignore the constant term $\|\nabla f(x)\|^2 / L^2$ and multiply by $L/2$ without changing the minimization problem.

This results in the following algorithm.

Projected gradient method
Given $x_0, L, k = 0$
Repeat
$\quad x_{k+1} = P_C(x_k - \frac{1}{L} \nabla f(x_k))$
$\quad k \leftarrow k + 1$