1 Optimization methods

1.1 Convex functions

Proposition: Let f be a C^1 function. Then f is convex if and only if dom$f$ is convex and ∇f is monotone,

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq 0.$$

Proof. $\Rightarrow$ Use the first order convexity criterion.
$\Leftarrow$ Consider $g(t) = f(x + t(y - x)$. Integrate $g'(t)$. (Try it!). □

Proposition: Let f be a convex C^1 function. Then the following are equivalent.

1. $\nabla f(x)$ is Lipschitz: there exists $L > 0$ such that

$$||\nabla f(x) - \nabla f(y)|| \leq L||x - y||, \ \forall x, y \in \text{dom} f$$

2. $g(x) := \frac{L}{2}||x||^2 - f(x)$ is convex.

3. Quadratic upper bound

$$f(y) \leq f(x) + \langle \nabla f(x), y - x \rangle + \frac{L}{2}||y - x||^2, \forall x, y \in \text{dom} f$$

Proof. (a)$\Rightarrow$ (b): Note that $\nabla g(x) = Lx - \nabla f(x)$.

$$\begin{aligned} \langle \nabla g(x) - \nabla g(y), x - y \rangle &= \langle L(x - y) - (\nabla f(x) - \nabla f(y), x - y) \rangle \\ &= L||x - y||^2 - \langle \nabla f(x) - \nabla f(y), x - y \rangle \\ &\geq L||x - y||^2 - ||\nabla f(x) - \nabla f(y)||||x - y|| \\ &\geq L||x - y||^2 - L||x - y||^2 \geq 0 \end{aligned}$$

Hence g is convex.
(b)$\Rightarrow$(c) Since g is convex, $g(y) \geq g(x) + \langle \nabla g(x), y - x \rangle$
So $\frac{L}{2}||y||^2 - f(y) \geq \frac{L}{2}||x||^2 - f(x) + \langle Lx - \nabla f(x), y - x \rangle$.
Hence $f(y) \leq f(x) + \langle \nabla f(x), y - x \rangle + \frac{L}{2}||y - x||^2$.
(c)$\Rightarrow$(a) $f(y) \leq f(x) + \langle \nabla f(x), y - x \rangle + \frac{L}{2}||y - x||^2$
$f(x) \leq f(y) + \langle \nabla f(y), x - y \rangle + \frac{L}{2}||x - y||^2$
Adding the two inequalities, we get,

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \leq L||x - y||^2$$

We need to show that $\langle \nabla f(x) - \nabla f(y), x - y\rangle \geq \frac{1}{L}||\nabla f(x) - \nabla f(y)||^2$.
Consider the function $\phi_x(z) := f(z) - \langle \nabla f(x), z\rangle$.
ϕ_x is convex and $\nabla \phi_x(z) = \nabla f(z) - \nabla f(x)$.
Since, $f(z) \leq f(y) + \langle \nabla f(y), z - y\rangle + \frac{L}{2}||z - y||^2$, we have

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

That is

$$\phi_x(z) \leq \phi_x(y) + \langle \nabla \phi_x(y), z - y\rangle + \frac{L}{2}||z - y||^2$$

We minimized both sides over z. The left hand side is minimized at $z = x$. The right hand side is minimized at $z = -\frac{1}{L}\nabla \phi_x(y) + y$. Hence,

$$\begin{aligned} f(x) - \langle \nabla f(x), x\rangle = \phi_x(x) &\leq \phi_x(y) + \langle \nabla \phi_x(y), -\frac{1}{L}\nabla \phi_x(y)\rangle + \frac{L}{2}||\frac{1}{L}\nabla \phi_x(y)||^2 \\ &= f(y) - \langle \nabla f(x), y\rangle - \frac{1}{2L}||\nabla f(y) - \nabla f(x)||^2 \end{aligned}$$

So

$$f(y) - f(x) - \langle \nabla f(x), y - x\rangle \geq \frac{1}{2L}\|\nabla f(y) - \nabla f(x)\|^2$$

Interchange the role of x, y, we get

$$f(x) - f(y) - \langle \nabla f(y), x - y\rangle \geq \frac{1}{2L}\|\nabla f(y) - \nabla f(x)\|^2$$

Adding the two inequalities, we get

$$\langle \nabla f(x) - \nabla f(y), x - y\rangle \geq \frac{1}{L}||\nabla f(x) - \nabla f(y)||^2$$

□

Proposition: Suppose f is a convex C^1 function. Suppose $\nabla f(x)$ is Lipschitz continuous with parameter L. Suppose x^* is a global minimum of f. Then

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

Proof. From quadratic upper bound, we get

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

But x^* is a global minimum, so $\nabla f(x^*) = 0$.
Now, consider $\inf_y (f(y) + \langle \nabla f(x) + y - x \rangle + \frac{L}{2}||y-x||^2$. It is minimized at $y = x - \frac{\nabla f(x)}{L}$.
Hence

$$f(x^*) = \inf_y f(y) \leq \inf_y \left(f(x) + \langle \nabla f(x), y - x \rangle + \frac{L}{2} \|y - x\|^2 \right) = f(x) - \frac{1}{2L} \|\nabla f(x)\|^2$$

□

1.2 Descent Methods

Consider the following minimization problem

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

where f is a convex differentiable function.

A general optimization algorithm is of the following form:
Choose initial point x^0 and repeat
$x^{k+1} = x^k + t_k d^k, \; k = 0, 1, ...$

What should we choose for d^k? What should we choose for t_k?
For the first question, we want d^k to be a descent direction, that is

$$\langle d^k, \nabla f(x^k) \rangle \leq 0$$

Since f is convex, if $\langle \nabla f(x^k), t_k d^k \rangle > 0$, then $f(x^{k+1}) > f(x^k)$. Hence in order to for the function value to descent, we must have

$$\langle d^k, \nabla f(x^k) \rangle \leq 0$$

As for the second question, there are mainlty three ways to select t_k.
Fixed step size: t_k is constant.
Exact line search

$$t_k = \text{argmin}_{s \geq 0} f(x + s d_k)$$

Backtracking line search: Choose $0 < \beta < 1$, initialize $t_k = 1$; take $t_k := \beta t_k$ until

$$f\left(x - t_k \nabla f(x)\right) < f(x) - \frac{1}{2} t_k \|\nabla f(x)\|^2$$

1.3 Gradient descent

In gradient descent, we choose d_k to be $\nabla f(x_k)$. So $x^{k+1} = x^k - t_k \nabla f\left(x^k\right)$.

Proposition: Suppose f is a convex C^1 function and $\nabla f(x)$ is Lipschitz with parameter L. If the step size $t \leq \frac{1}{L}$, then the fixed size gradient descent satisfies

$$f(x^k) - f(x^*) \leq \frac{1}{2kt} \left\|x^0 - x^*\right\|^2$$

Proof. Let $x^+ := x - t\nabla f(x)$. Then using quadratic upper bound, we have,

$$f\left(x^+\right) \leq f(x) + \left(-t + \frac{Lt^2}{2}\right) \|\nabla f(x)\|^2 \leq f(x) - \frac{t}{2}\|\nabla f(x)\|^2$$

Since f is convex, $f\left(x^*\right) \geq f(x) + \langle \nabla f(x), x^* - x\rangle$. Then

$$\begin{aligned} f\left(x^+\right) &\leq f(x) - \frac{t}{2}\|\nabla f(x)\|^2 \\ &\leq f^* + \langle \nabla f(x), x - x^*\rangle - \frac{t}{2}\|\nabla f(x)\|^2 \\ &= f^* + \frac{1}{2t}\left(\|x - x^*\|^2 - \|x - x^* - t\nabla f(x)\|^2\right) \\ &= f^* + \frac{1}{2t}\left(\|x - x^*\|^2 - \left\|x^+ - x^*\right\|^2\right) \end{aligned}$$

Summing the above, we get

$$\begin{aligned} \sum_{i=1}^{k}\left(f\left(x^i\right) - f^*\right) &\leq \frac{1}{2t}\sum_{i=1}^{k}\left(\left\|x^{i-1} - x^*\right\|^2 - \left\|x^i - x^*\right\|^2\right) \\ &= \frac{1}{2t}\left(\left\|x^0 - x^*\right\|^2 - \left\|x^k - x^*\right\|^2\right) \\ &\leq \frac{1}{2t}\left\|x^0 - x^*\right\|^2 \end{aligned}$$

But $f(x^i)$ is decreasing, hence

$$f\left(x^k\right) - f^* \leq \frac{1}{k}\sum_{i=1}^{k}\left(f\left(x^i\right) - f^*\right) \leq \frac{1}{2kt}\left\|x^0 - x^*\right\|^2$$

□