Conclusion

Ernest K. Ryu
Seoul National University

Mathematical and Numerical Optimization
Fall 2020

Last edited: 12/10/2020

Conclusion

The goal of this class was to present a unified analysis of convex optimization algorithms through the abstraction of monotone operators.

Optimization is useful. (It is *applied* math.) However, I personally think this subject also has a certain beauty, and I wanted to share it with you.

There are a few additional topics, 4 chapters, we were unable to cover. The following slides briefly summarize them.

Asynchronous coordinate update methods

Asynchronous coordinate-update fixed-point iteration (AC-FPI):

```
// p agents run the while loop asynchronously
// x is a vector stored in shared memory
WHILE (not converged) {
  1. Select i from Uniform{1,2,...,m}
  2. Read x
  3. Compute s[i] = eta*S[i](x)
  4. Exclusively read x[i] and
                  write x[i] = x[i] - s[i]
}
```

Exclusive access through atomic operations or mutex.

With $\mathbb{T} = \mathbb{I} - \theta\mathbb{S}$, mathematically model algorithm as:

$$x^{k+1} = x^k - \eta \mathbb{S}_{i(k)} x^{k-\boldsymbol{d}(k)}$$

Under suitable assumptions, converges almost surely to a solution.

ADMM-type methods

Consider the primal problem

$$\begin{array}{ll} \underset{x\in\mathbb{R}^p,\, y\in\mathbb{R}^q}{\text{minimize}} & f_1(x)+f_2(x)+g_1(y)+g_2(y) \\ \text{subject to} & Ax+By=c, \end{array}$$

generated by the Lagrangian

$$\mathbf{L}(x,y,u)=f_1(x)+f_2(x)+g_1(y)+g_2(y)+\langle u, Ax+By-c\rangle.$$

Function-linearized proximal alternating direction method of multipliers (FLiP-ADMM):

$$x^{k+1}\in\underset{x\in\mathbb{R}^p}{\operatorname{argmin}}\left\{f_1(x)+\langle\nabla f_2(x^k)+A^\intercal u^k, x\rangle+\frac{\rho}{2}\|Ax+By^k-c\|^2+\frac{1}{2}\|x-x^k\|_P^2\right\}$$

$$y^{k+1}\in\underset{y\in\mathbb{R}^q}{\operatorname{argmin}}\left\{g_1(y)+\langle\nabla g_2(y^k)+B^\intercal u^k, y\rangle+\frac{\rho}{2}\|Ax^{k+1}+By-c\|^2+\frac{1}{2}\|y-y^k\|_Q^2\right\}$$

$$u^{k+1}=u^k+\varphi\rho(Ax^{k+1}+By^{k+1}-c),$$

where $\rho>0$, $\varphi>0$, $P\in\mathbb{R}^{p\times p}$ and $P\succeq 0$, and $Q\in\mathbb{R}^{q\times q}$ and $Q\succeq 0$. Converges under suitable conditions.

Stochastic optimization

Consider

$$\underset{x \in \mathbb{R}^n}{\text{find}} \quad 0 \in \left(\frac{1}{N} \sum_{i=1}^{N} \mathbb{A}_i + \mathbb{B} \right) x.$$

Stochastic forward-backward method (SFB):

$$x^{k+1} \in \mathbb{J}_{\alpha_k \mathbb{B}} (\mathbb{I} - \alpha_k \mathbb{A}_{i(k)}) x^k,$$

where $\alpha_k > 0$ and $i(k) \in \{1, \dots, N\}$ independently uniformly at random.

The famous stochastic gradient descent (SGD) is an instance of SFB.

Under suitable assumptions, converges almost surely to a solution.

Acceleration: Accelerated gradient method

Consider

$$\underset{x\in\mathbb{R}^n}{\text{minimize}} \quad f(x),$$

where f is convex and L-smooth. The method

$$\begin{aligned} x^{k+1} &= y^k - \frac{1}{L}\nabla f(y^k) \\ y^{k+1} &= x^{k+1} + \frac{k-1}{k+2}(x^{k+1} - x^k), \end{aligned}$$

where $x^0 = y^0$, is Nesterov's accelerated gradient method (AGM).

Theorem 1.

Assume the convex, L-smooth function f has a minimizer $x^\star$. Then AGM converges with the rate

$$f(x^k) - f(x^\star) \leq \frac{2L\|x^0 - x^\star\|^2}{k^2}.$$

Acceleration: Accelerated proximal point

Consider

$$\underset{x \in \mathbb{R}^n}{\text{find}} \quad 0 \in \mathbb{A}x,$$

where $\mathbb{A}$ is maximal monotone. The method

$$\begin{aligned} y^{k+1} &= \mathbb{J}_{\mathbb{A}} x^k \\ x^{k+1} &= y^{k+1} + \frac{k}{k+2}(y^{k+1} - y^k) - \frac{k}{k+2}(y^k - x^{k-1}), \end{aligned}$$

where $y^0 = x^0$, is the accelerated proximal point method (APPM).

Theorem 2.

Assume the maximal monotone operator $\mathbb{A}$ has a zero $x^\star$. Then APPM converges with the rate

$$\|x^{k-1} - \mathbb{J}_{\mathbb{A}} x^{k-1}\|^2 \leq \frac{\|x^0 - x^\star\|^2}{k^2}.$$