EECS 227C Convex Optimization and Approximation Lecture 19 - 3/29/2022

Lecture 19: Potential Reduction Interior Point Method

Lecturer: Yanjun Han *Scribe: Homer Walke, Gaofeng Su*

In this lecture we cover conic duality, Karmarkar's algorithm, and the primal-dual potential reduction algorithm. They are based on [1, Chapter 5-7].

1 Conic Duality

First we define a conic program.

Conic Program

$$\begin{aligned} \min \quad & c^T x \\ \text{s.t.} \quad & x \in K \cap \{b + L\} \end{aligned}$$

where K is a convex cone ($x \in K, t \geq 0 \Rightarrow tx \in K$). Some examples of convex cones include: $K = \mathbb{R}^n_+$, $K = \{x \in \mathbb{R}^n \mid x_n^2 \geq x_1^2, \ldots, x_{n-1}^2\}$, and $K = \{x \in \mathbb{R}^{n \times m} \mid x \succeq 0\}$.

Note that general convex programming can be represented as conic programming.

Convex Program		**Conic Program**
$\min \quad c^T x$	$=$	$\max \quad (c, 0)^T (x, t)$
$\text{s.t.} \quad x \in G$		$\text{s.t.} \quad (x, t) \in \{t > 0, t^{-1} x \in G\}, t = 1$

We can also define a dual conic program.

Primal Problem	**Dual Problem**
$\min \quad c^T x$	$\max \quad b^T y$
$\text{s.t.} \quad x \in K \cap \{b + L\}$	$\text{s.t.} \quad y \in K^* \cap \{c + L^\perp\}$

Here, K^* is the dual cone of K. The dual cone is defined as $K^* = \{y \in \mathbb{R}^n \mid \langle x, y \rangle \geq 0 \ \forall x \in K\}$. There are two versions of duality, weak and strong.

Weak Duality: For feasible (x, y), then

$$c^T x + b^T y - c^T b = -(x - b)^T (y - c) + x^T y = x^T y \geq 0.$$

Strong Duality: If the primal problem is strongly feasible and lower bounded, then the dual problem is solvable. Moreover,

$$P^* + D^* - c^T b = 0$$

1.1 Log-homogeneous barrier

Definition 1 (Log-homogeneous barrier). *F is a θ-logarithmically homogeneous self-concordant barrier for K if and only if F is a self-concordant barrier for K and $F(tx) = F(x) - \theta \log t$.*

Proposition 2. *We claim $\nabla F(x) = -\nabla^2 F(x) x$ and $x^T \nabla^2 F(x) x \equiv \theta$ and F is θ-self-concordant.*

Proof Note that

$$\nabla F(tx) = t^{-1}\nabla F(x)$$

Taking the derivative of this expression with respect to t and setting $t = 1$, we have

$$\nabla^2 F(x)x = -\nabla F(x)$$

Then taking the derivative of $F(tx) = F(x) - \theta \log t$ with respect to t and setting $t = 1$, we have

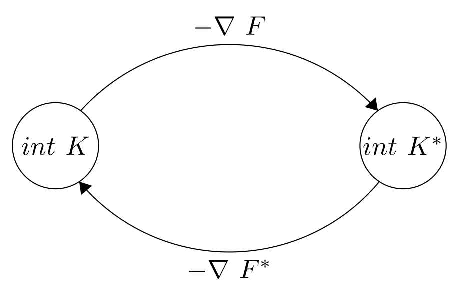
$$-\theta = \nabla F(x)^T x = -x^T \nabla^2 F(x)x$$

Lastly we have

$$|\langle \nabla F(x), h\rangle| = |h^T \nabla^2 F(x)x| \leq ||x||_x^{1/2}||h||_2^{1/2} = \sqrt{\theta}\sqrt{h^T \nabla F(x)h}$$

□

Proposition 3. *We claim that $F^*(y) = F^*(-y) = \sup_{x\in K}(-x^T y - F(x))$ is a θ-logarithmically homogeneous self-concordant barrier for K^*.*
Also,



Moreover,

$$F(x) + F^*(-y) + \theta \log(x^T y) \geq \theta \log \theta - \theta$$

with equality if and only if $y = -t\nabla F(x)$ for some $t \geq 0$.

Here we prove the last part of the proposition.
Proof First we show the $\Leftarrow$ direction. If $y = -t\nabla F(x)$, then $x^T y = -t\nabla F(x)^T x = t\theta$. So

$$\begin{aligned} F(x) + F^*(-y) + \theta \log(x^T y) &= F(x) + F^*(\nabla F(x)) - \theta \log t + \theta \log(t\theta) \\ &= \theta \log \theta + \nabla F(x)^T x = \theta \log \theta - \theta \end{aligned}$$

Now we show the $\Rightarrow$ direction. Minimize $F(z)$ over $\{x^T y = z^T y\}$ then $\nabla F(x) = -ty$ with $t \geq 0$. and the objective value is $\theta \log \theta - \theta$. □

2 Karmarkar's Algorithm

We want to solve the primal problem:

$$\begin{aligned} \min \quad & c^T x \\ \text{s.t.} \quad & x \in K \cap \text{Ł}, \quad e^T x = 1. \end{aligned}$$

We make the following assumptions:

1. The feasible set is bounded.
2. There is a known feasible solution.
3. There is a known barrier F.
4. There is a known optimal objective value c^*.

Under these assumptions the following problem has objective value 0.

$$\begin{aligned} \min \quad & (c - c^* e)^T x \triangleq r^T x \\ \text{s.t.} \quad & x \in K \cap L, \quad e^T x = 1. \end{aligned}$$

The idea of the algorithm is to minimize the Karmarkar potential:

$$v(x) = F(x) + \theta \log(\sigma^T x) \quad \text{s.t. } x \in L$$

Note that $v(tx) = v(x)$. We know that

$$v(x) \leq -A \Rightarrow \sigma^T x \leq C e^{-A/\theta}$$

The algorithm has the following steps:

1. $v(x) \leq v_{x_t}(x) = F(x) + \theta \frac{\sigma^T x}{\sigma^T x_t} - \theta + \theta \log(\sigma^T x_t)$
2. Then we caculate the Newton direction:
$$\begin{aligned} e_t = \arg\min_h & \, h^T \nabla v_{x_t}(x_t) + \frac{1}{2} h^T \nabla^2 v_{x_t}(x_t) h \\ \text{s.t. } & h \in Ł, \quad h^T \nabla F(x_t) = 0. \end{aligned}$$
3. Then we use the damped Newton method:
$$x' = x_t + \frac{1}{1+w} e_t \quad w = \sqrt{-e_t^T \nabla v_{x_t}(x_t)}$$
4. Set x'' to be any $v(x'') \leq v(x')$ and $x_{t+1} = \frac{x''}{e^T x''}$

Theorem 4. $v(x_T) - v(x_{t+1}) \geq \frac{1}{3} - \log \frac{4}{3} \geq 0$

Analysis:

$$\begin{aligned} v(x_T) - v(x_{t+1}) &= v(x_T) - v(x'') \\ &\geq v(x_t) - v(x') \\ &\geq v_{x_t}(x_t) - v_{x_t}(x') \\ &\geq w - \log(1+w) \end{aligned}$$

It then suffices to prove that the Newton decrement $w \geq \frac{1}{3}$.

Why do we use potential reduction? While the objective is the same, we can make much greater progress in practice.

What if we don't know c^*? First find a lower bound $c_t \leq c^*$. Then run the procedure with c_t.

1. If $w_t \geq \frac{1}{3}$, continue.
2. If $w_t < \frac{1}{3}$, find (using grid search) the smallest $c_{t+1} > c_t$ such that $w_t \geq \frac{1}{3}$ with c_{t+1}. Start using c_{t+1} and continue.

3 Primal-dual Potential Reduction Algorithm

Given the primal and dual problem setup as follows:

Primal Problem	**Dual Problem**
$\min \quad c^T x$	$\max \quad b^T y$
s.t. $\quad x \in K \cap \{b + L\}$	s.t. $\quad y \in K^* \cap \{c + L^{\perp}\}$

The idea of primal-dual potential reduction algorithm is to minimize the potential function:

$$v(x, y) = F(x) + F^{+}(y) + (\theta + \mu) \log(x^T y)$$

Because,

$$\text{if} \quad v(x, y) \leq -A \Rightarrow x^T y \leq exp(\frac{-A + \theta(\log\theta - 1)}{\mu})$$

note: $F(x) + F^{+}(y) + \mu \log(x^T y) \geq \theta \log \theta - \theta$ (see Prop 3)

3.1 Algorithm Idea

The first idea is called **primal update**, where we fix y_t and locally linearize $v(x, y_t)$ around $x = x_t$,

$$v(x, y_t) \leq v_t(x) = F(x) + (\theta + \mu) \frac{x^T y_t}{x_t^T y_t} + const(y_t)$$

and let

$$\begin{aligned} e_t = &\arg\min_h h^T \nabla v_t(x_t) + \frac{1}{2} h^T \nabla^2 v_t(x_t) h \\ &\text{s.t.} \quad h \in L \end{aligned}$$

Then the updating rule is:

$$\begin{aligned} w_t &= \sqrt{-e_t^T \nabla v_t(x_t)} \\ x_{t+1} &= x_t + \frac{1}{1 + w_t} e_t \end{aligned}$$

Analysis:

$$v(x_{t+1}, y_t) - v(x_t, y_t) \leq \log(1 + w_t) - w_t \leq 0$$

Problem: w_t could be small (or even 0), which means we are making small (or even zero) progress.
Fix: when w_t is small, we can make great progress for the **dual update**, which we will define later.
Idea: suppose that $w_t = 0$, then

$$\nabla v_t(x_t) = \nabla F(x_t) + (\theta + \mu) \frac{y_t}{x_t^T y_t} \in L^{\perp} \Leftrightarrow -\frac{x_t^T y_t}{\theta + \mu} \nabla F(x_t) \in y_t + L^{\perp} = c + L^{\perp}$$

Try

$$y_{t+1} = -\frac{x_t^T y_t}{\theta + \mu} \nabla F(x_t)$$

then

$$
\begin{aligned}
v(x_t, y_t) - v(x_t, y_{t+1}) =& \underbrace{(F(x_t) + F^{+}(y_t) + \theta \log(x_t^T y_t)) - (F(x_t) + F^{+}(y_{t+1}) + \theta \log(x_t^T y_{t+1}))}_{\geq 0} \\
&+ \mu \log(x_t^T y_t)) - \mu \log(x_t^T y_{t+1})) \\
\geq& \mu \log(\frac{\theta + \mu}{-\nabla F(x_t)^T x_t}) = \mu \log(1 + \frac{\theta}{\mu})
\end{aligned}
$$

Dual update:

$$
y_{t+1} = -\frac{x_t^T y_t}{\theta + \mu}(\nabla F(x_t) + \nabla^2 F(x_t) e_t)
$$

Theorem 5. *Given $y_{t+1} \in c + L^{\perp}$, if $w_t < 1$ and $y_{t+1} \in K^*$, then $v(x_t, y_t) - v(x_t, y_{t+1}) \geq \mu \log \frac{\theta+\mu}{\theta+w_t\sqrt{\theta}} + w_t + \log(1 - w_t)$*

Analysis:

$$
\text{defn. of } e_t \Rightarrow \nabla F(x_t) + \frac{\theta + \mu}{x_t^T y_t} y_t + \nabla^2 F(x_t) e_t \in L^{\perp} \Rightarrow y_{t+1} - y_t \in L^{\perp} \Rightarrow y_{t+1} \in c + L^{\perp}
$$

Choose μ $\sqrt{\theta}$ to guarantee constant progress, and the number of iteration $\leq \tilde{O}(\sqrt{\theta} \log \frac{1}{\epsilon})$

3.2 Primal-dual potential-reduction algorithm

1. start from feasible solution (x_p, y_p)
2. at t^{th} iteration, run primal update $\rightarrow (x_{t+1}, y_t)$ and dual update $\rightarrow (x_t, y_{t+1})$
3. if y_{t+1} is not strictly dual feasible, choose (x_{t+1}, y_t), otherwise choose the one with smaller $v(x, y)$

References

[1] A. Nemirovski, "Interior point polynomial methods in convex programming," https://www2.isye.gatech.edu/~nemirovs/Lect_IPM.pdf, accessed: 2022-04-05.