Nesterov's estimate sequence:
1. What is it and how to construct one

Andersen Ang

Department of Combinatorics and Optimization,
University of Waterloo, Waterloo, Canada

ms**x**ang@uwaterloo.ca, where $\mathbf{x} = \lfloor \pi \rfloor$
Homepage: angms.science

First draft: November 21, 2021 Last update: November 27, 2021

Minimization Problem

- Consider

$$\min_{\mathbf{x}} f(\mathbf{x})$$

where f is convex and smooth.

- **Definition** A pair of sequences $\{\phi_k(\mathbf{x}), \lambda_k\}_{k=0}^{\infty}$, $\lambda_k \geq 0$ is called an estimate sequence of $f(\mathbf{x})$ if
  - $\lambda_k \to 0$ as k increases
  - $\phi_k(\mathbf{x}) \leq (1 - \lambda_k) f(\mathbf{x}) + \lambda_k \phi_0(\mathbf{x})$ for all $k \geq 0$

Why estimate sequence: it forms an upper bound to $f_k - f^*$

- If for some sequence $\{\mathbf{x}_k\}$ we have

$$f(\mathbf{x}_k) \leq \phi_k^* := \min_{\mathbf{x}\in\mathbb{R}^n} \phi_k(\mathbf{x}), \tag{3}$$

  then

$$f(\mathbf{x}_k) - f^* \leq \lambda_k \Big(\phi_0(\mathbf{x}^*) - f^*\Big) \to 0.$$

- This explains why the estimate sequence $\{\phi_k(\mathbf{x}), \lambda_k\}_{k=0}^{\infty}$, $\lambda_k \geq 0$ could be useful:
  - It forms an upper bound the of the cost optimality gap $f(\mathbf{x}_k) - f^*$
  - This upper bound converges to 0, due to the definition of $\lambda_k \to 0$ as k increases. (Noting that $\phi_0(\mathbf{x}^*) - f^*$ is a constant.)
  - For any $\{\mathbf{x}_k\}$, if (3) is satisfied, then the convergence rate of $\{\mathbf{x}_k\}$ can be derived from the convergence rate of sequence $\{\lambda_k\}$

The proof

- **Definition** The sequences $\{\phi_k(\mathbf{x}), \lambda_k\}_{k=0}^{\infty}$, $\lambda_k \geq 0$ is an estimate sequence of $f(\mathbf{x})$ if
  - $\lambda_k \to 0$ as k increases (1)
  - $\phi_k(\mathbf{x}) \leq (1 - \lambda_k) f(\mathbf{x}) + \lambda_k \phi_0(\mathbf{x})$ for all $k \geq 0$ (2)

- **Lemma**: If for some sequence $\{\mathbf{x}_k\}$ we have
$$f(\mathbf{x}_k) \leq \phi_k^* := \min_{\mathbf{x} \in \mathbb{R}^n} \phi_k(\mathbf{x}), \tag{3}$$
then $f(\mathbf{x}_k) - f^* \leq \lambda_k \Big(\phi_0(\mathbf{x}^*) - f^*\Big) \to 0$.

- **The proof:**
$$f(\mathbf{x}_k) \overset{(3)}{\leq} \phi_k^* \overset{(3)}{=} \min_{\mathbf{x} \in \mathbb{R}^n} \phi_k(\mathbf{x}) \overset{(2)}{\leq} \min_{\mathbf{x} \in \mathbb{R}^n} (1 - \lambda_k) f(\mathbf{x}) + \lambda_k \phi_0(\mathbf{x}) \leq (1 - \lambda_k) f(\mathbf{x}^*) + \lambda_k \phi_0(\mathbf{x}^*).$$
This gives
$$f(\mathbf{x}_k) - f^* \leq \lambda_k \Big(\phi_0(\mathbf{x}^*) - f^*\Big) \overset{(1)}{\to} 0.$$

A way to form an estimate sequence

- Now we know that estimate sequence can be used to derive convergence rate of $\{\mathbf{x}_k\}$, then how to form an estimate sequence?

- **Lemma** Assume 5 conditions:
  1. f is μ-strongly convex and L-smooth
  2. $\phi_0(\mathbf{x})$ is an arbitrary function in $\mathbb{R}^n$
  3. $\{\mathbf{y}_k\}_{k=0}^{\infty}$ is an arbitrary sequence in $\mathbb{R}^n$
  4. $\{\alpha_k\}_{k=0}^{\infty} : \alpha_k \in]0,1[\,, \displaystyle\sum_{k=0}^{\infty} \alpha_k = \infty$
  5. $\lambda_0 = 1$

  Then the pair of sequences $\{\phi_k(\mathbf{x}), \lambda_k\}_{k=0}^{\infty}$, $\lambda_k \geq 0$ defined as

$$\begin{aligned} \lambda_{k+1} &= (1-\alpha_k)\lambda_k, \\ \phi_{k+1}(\mathbf{x}) &= (1-\alpha_k)\phi_k(\mathbf{x}) + \alpha_k\Big(f(\mathbf{y}_k) + \langle \nabla f(\mathbf{y}_k), \mathbf{x}-\mathbf{y}_k\rangle + \tfrac{\mu}{2}\|\mathbf{x}-\mathbf{y}_k\|_2^2\Big) \end{aligned}$$

  is an estimate sequence of $f(\mathbf{x})$.

The proof

- To prove that

$$\begin{aligned} \lambda_{k+1} &= (1-\alpha_k)\lambda_k, \\ \phi_{k+1}(\mathbf{x}) &= (1-\alpha_k)\phi_k(\mathbf{x}) + \alpha_k\Big(f(\mathbf{y}_k) + \langle \nabla f(\mathbf{y}_k), \mathbf{x}-\mathbf{y}_k\rangle + \tfrac{\mu}{2}\|\mathbf{x}-\mathbf{y}_k\|_2^2\Big) \end{aligned}$$

  is an estimate sequence of $f(\mathbf{x})$, we need to show that
  - $\{\lambda_k\}$ defined in this way converges to 0
  - $\phi_k(\mathbf{x})$ defined in this way satisfies $\phi_k \le \phi_k(\mathbf{x}) \le (1-\lambda_k)f(\mathbf{x}) + \lambda_k\phi_0(\mathbf{x})$ for all k

- We will first prove $\lambda_k \to 0$

- Then we show $\phi_k \le \phi_k(\mathbf{x}) \le (1-\lambda_k)f(\mathbf{x}) + \lambda_k\phi_0(\mathbf{x})$ by induction.

Showing $\lambda_k \to 0$

- By definition, $\lambda_{k+1} = (1-\alpha_k)\lambda_k$ with $\alpha_k \in]0,1[$ and $\lambda_0 = 1$, so the sequence $\{\lambda_k\}$ is monotone decreasing.
- The sequence $\{\lambda_k\}$ is bounded below by 0.
- Monotone decreasing + bounded below $\implies$ $\{\lambda_k\}$ has a limit, let it be c.
- We now show $c = 0$ by contradiction.
  Suppose $c > 0$, i.e., the sequence $\{\lambda_k\}$ converges to a positive number. Now consider $\lambda_k - \lambda_{k+1} = \lambda_k - (1-\alpha_k)\lambda_k = \alpha_k\lambda_k$. Notice that it forms a telescoping sum, summing it from 0 to k gives
  $$\lambda_0 - \lambda_{k+1} = \sum_{i=0}^{k} \alpha_i \lambda_i \geq \sum_{i=0}^{k} \alpha_i c = c\sum_{i=0}^{k} \alpha_i \qquad (*)$$
  where the $\geq$ is based on the fact that we assume $\{\lambda_k\}$ converges to c, then all $\lambda_k \geq c$ for all k.
- Now we have $\lambda_0 - \lambda_{k+1} \geq c\sum_{i=0}^{k} \alpha_i$. Take limit $k \to \infty$ gives $\lambda_0 - c \geq c\sum_{i=0}^{\infty} \alpha_i$. By condition 4 $\sum_{i=0}^{\infty} \alpha_i = +\infty$ so $\lambda_0 - c \geq +\infty$, which is impossible, a contradiction, therefore $c = 0$.

Proof on ϕ_k

- Case $k = 0$. $\phi_0(\mathbf{x}) \leq (1 - \lambda_0) f(\mathbf{x}) + \lambda_0 \phi_0(\mathbf{x}) = \phi_0(\mathbf{x})$ by condition 5 that $\lambda_0 = 1$.
- Induction hypothesis: let $\phi_k \leq (1 - \lambda_k) f(\mathbf{x}) + \lambda_k \phi_0(\mathbf{x})$ is true for some $k \geq 0$.
- Case $k + 1$. By definition $\phi_{k+1}(\mathbf{x}) = (1 - \alpha_k)\phi_k(\mathbf{x}) + \alpha_k \Big(f(\mathbf{y}_k) + \langle \nabla f(\mathbf{y}_k), \mathbf{x} - \mathbf{y}_k \rangle + \frac{\mu}{2} \|\mathbf{x} - \mathbf{y}_k\|_2^2 \Big)$.
  By condition 1 that f is μ-strongly convex, we have

$$\begin{aligned} \phi_{k+1}(\mathbf{x}) &= (1 - \alpha_k)\phi_k(\mathbf{x}) + \alpha_k \Big(f(\mathbf{y}_k) + \langle \nabla f(\mathbf{y}_k), \mathbf{x} - \mathbf{y}_k \rangle + \frac{\mu}{2} \|\mathbf{x} - \mathbf{y}_k\|_2^2 \Big) \\ &\leq (1 - \alpha_k)\phi_k(\mathbf{x}) + \alpha_k f(\mathbf{x}) \end{aligned}$$

  So we now have

$$\phi_{k+1}(\mathbf{x}) \leq (1 - \alpha_k)\phi_k(\mathbf{x}) + \alpha_k f(\mathbf{x})$$

- See slide 8 here for the details of strongly convex function.

- continue

$$
\begin{aligned}
\phi_{k+1}(\mathbf{x}) &\leq (1-\alpha_k)\phi_k(\mathbf{x}) + \alpha_k f(\mathbf{x}) \\
&\overset{\text{tricky}}{=} (1-\alpha_k)\Big(\phi_k(\mathbf{x}) + \underline{(1-\lambda_k)f(\mathbf{x})} - (1-\lambda_k)f(\mathbf{x})\Big) + \alpha_k f(\mathbf{x}) \\
&= (1-\alpha_k)\Big(\underbrace{\phi_k(\mathbf{x}) - (1-\lambda_k)f(\mathbf{x})}_{\leq \lambda_k \phi_0(\mathbf{x})}\Big) + \underbrace{(1-\alpha_k)\underline{(1-\lambda_k)f(\mathbf{x})}}_{=\Big((1-\alpha)-(1-\alpha)\lambda\Big)f} + \alpha_k f(\mathbf{x}) \\
&= (1-\alpha_k)\lambda_k\phi_0(\mathbf{x}) + \Big(1-(1-\alpha_k)\lambda_k\Big)f(\mathbf{x}).
\end{aligned}
$$

- By construct $\lambda_{k+1} = (1-\alpha_k)\lambda_k$ so we have

$$
\phi_{k+1}(\mathbf{x}) \leq \lambda_{k+1}\phi_0(\mathbf{x}) + (1-\lambda_{k+1})f(\mathbf{x}).
$$

- By mathematical induction, the pair of sequences $\{\phi_k(\mathbf{x}), \lambda_k\}_{k=0}^{\infty}$, $\lambda_k \geq 0$ defined as

  $\lambda_{k+1} = (1-\alpha_k)\lambda_{k+1}$, $\phi_{k+1}(\mathbf{x}) = (1-\alpha_k)\phi_k(\mathbf{x}) + \alpha_k\Big(f(\mathbf{y}_k) + \langle \nabla f(\mathbf{y}_k), \mathbf{x}-\mathbf{y}_k\rangle + \frac{\mu}{2}\|\mathbf{x}-\mathbf{y}_k\|_2^2\Big)$

  is an estimate sequence of $f(\mathbf{x})$. $\square$

Last page - summary

- Nesterov's estimate sequence: what is it and how to construct one

Reference:

- Yurii Nesterov, Introductory lectures on convex optimization: a basic course, Kluwer Academic Publishers, 2003.

End of document