Robustness of Stochastic Extrapolated Model-based methods

We carry out a brief analysis of the robustness for extrapolated SMOD in the convex case. For the sake of asymptotic analysis, stepsize parameter γ_k in extra-SMOD is now indexed by k rather than being fixed at certain value.

We want to show

Theorem: Under assumption A8, in extra-SMOD we have

$$\mathbb{E}[\|x^* - x^{K+1}\|^2] \le \|x^* - x^0\|^2 + \beta(1-\beta)\|x^1 - x^0\|^2 + \sum_{k=0}^{K} \frac{2}{\theta^2\gamma_k^2}\mathbb{E}[\|f'(x^*, \xi_k)\|^2],$$

This result guarantees that with probability one, the iterates are bounded.

The result shows that extra-SMOD under Assumption A8 will have bounded iterates, Our result is an extension of [AD2019] which applies to stochastic (approximate) proximal point.

Our proof is a simple extension of the complexity analysis of extra-SMOD in the convex setting. Recall that by **A8** (Appendix C, Line 597) in convex case, we have that

$$\begin{aligned} f_{x^k}(x^{k+1}, B_k) &\ge f(x^{k+1}, B_k) - \frac{\tau}{2}\|x^{k+1} - x^k\|^2 \\ -f_{x^k}(x, B_k) &\ge -f(x, B_k) \end{aligned}$$

and summation over the above two relations gives

$$f_{x^k}(x^{k+1}, B_k) - f_{x^k}(x, B_k) + \frac{\tau}{2}\|x^{k+1} - x^k\|^2 \ge f(x^{k+1}, B_k) - f(x, B_k).$$

Starting from eqn(70) (Appendix C, Line 623), we have

$$\begin{aligned} f(x^{k+1}, B_k) - f(x, B_k) &\le f_{x^k}(x^{k+1}, B_k) - f_{x^k}(x, B_k) + \frac{\tau}{2}\|x^{k+1} - x^k\|^2 \\ &\le \frac{\gamma_k}{2}\|x - y^k\|^2 - \frac{\gamma_k}{2}\|x - x^{k+1}\|^2 - \frac{\gamma_k}{2}\|y^k - x^{k+1}\|^2 + \frac{\tau}{2}\|x^{k+1} - x^k\|^2, \end{aligned}$$

which implies that

$$\frac{\gamma_k}{2}\|x - x^{k+1}\|^2 \le \frac{\gamma_k}{2}\|x - y^k\|^2 + \frac{\tau}{2}\|x^{k+1} - x^k\|^2 - \frac{\gamma_k}{2}\|y^k - x^{k+1}\|^2 - [f(x^{k+1}, B_k) - f(x, B_k)] \tag{1}$$

Also, by the convexity of $f(\cdot, B_k)$, we have, for any $\eta > 0$ that

$$\begin{aligned} & f(x^{k+1}, B_k) - f(x, B_k) \\ \ge\; & \langle f'(x, B_k), x^{k+1} - x\rangle \\ =\; & \langle f'(x, B_k), x^k - x\rangle + \langle f'(x, B_k), x^{k+1} - x^k\rangle \\ \ge\; & \langle f'(x, B_k), x^k - x\rangle - \|f'(x, B_k)\|\|x^{k+1} - x^k\| \\ \ge\; & \langle f'(x, B_k), x^k - x\rangle - \frac{1}{2\eta\gamma_k}\|f'(x, B_k)\|^2 - \frac{\eta\gamma_k}{2}\|x^{k+1} - x^k\|^2. \end{aligned} \tag{2}$$

We also recall that

$$\begin{aligned} x - y^k &= \theta(\hat{x} - z^k) \\ x - x^{k+1} &= \theta(\hat{x} - z^{k+1}) \end{aligned}$$

By combining the above three parts, we have

$$
\begin{aligned}
&\frac{\gamma_k\theta^2}{2}\|\hat{x}-z^{k+1}\|^2 \\
&= \frac{\gamma_k}{2}\|x-x^{k+1}\|^2 \\
&\le \frac{\gamma_k}{2}\|x-y^k\|^2 + \frac{\tau}{2}\|x^{k+1}-x^k\|^2 - \frac{\gamma_k}{2}\|y^k-x^{k+1}\|^2 - [f(x^{k+1},B_k)-f(x,B_k)] \\
&\le \frac{\gamma_k}{2}\|x-y^k\|^2 + \frac{\tau}{2}\|x^{k+1}-x^k\|^2 - \frac{\gamma_k}{2}\|y^k-x^{k+1}\|^2 + \frac{\eta\gamma_k}{2}\|x^{k+1}-x^k\|^2 - \langle f'(x,B_k), x^k-x\rangle + \frac{1}{2\eta\gamma_k}\|f'(x,B_k)\|^2 \\
&= \frac{\gamma_k\theta^2}{2}\|\hat{x}-z^k\|^2 + \left\{\frac{\tau+\eta\gamma_k}{2}\|x^{k+1}-x^k\|^2 - \frac{\gamma_k}{2}\|y^k-x^{k+1}\|^2\right\} - \langle f'(x,B_k), x^k-x\rangle + \frac{1}{2\eta\gamma_k}\|f'(x,B_k)\|^2,
\end{aligned}
\tag{3}
$$

where the first inequality is from (1) and the second inequality is by (2).

Then following Appendix C Line 628, we bound $\left\{\frac{\tau+\eta\gamma_k}{2}\|x^{k+1}-x^k\|^2 - \frac{\gamma_k}{2}\|y^k-x^{k+1}\|^2\right\}$ by

$$
\begin{aligned}
&\frac{\tau+\eta\gamma_k}{2}\|x^{k+1}-x^k\|^2 - \frac{\gamma_k}{2}\|y^k-x^{k+1}\|^2 \\
&\le \frac{\gamma_k\beta(1-\beta)}{2}\|x^k-x^{k-1}\|^2 - \frac{\gamma_k(1-\beta-\eta)-\tau}{2}\|x^{k+1}-x^k\|^2.
\end{aligned}
\tag{4}
$$

Now we take expectation and combine (3) with (4) to get, for any x that

$$
\begin{aligned}
&\frac{\gamma_k\theta^2}{2}\mathbb{E}_k[\|\hat{x}-z^{k+1}\|^2] \\
&\le \frac{\gamma_k\theta^2}{2}\|\hat{x}-z^k\|^2 - \mathbb{E}_k[\langle f'(x,B_k), x^k-x\rangle] + \frac{1}{2\eta\gamma_k}\mathbb{E}_k[\|f'(x,B_k)\|^2] + \frac{\gamma_k\beta(1-\beta)}{2}\|x^k-x^{k-1}\|^2 \\
&\quad - \frac{\gamma_k(1-\beta-\eta)-\tau}{2}\mathbb{E}_k[\|x^{k+1}-x^k\|^2]
\end{aligned}
$$

and dividing both sides by $(\gamma_k\theta^2/2)$ gives

$$
\begin{aligned}
&\mathbb{E}_k[\|\hat{x}-z^{k+1}\|^2] \\
&\le \|\hat{x}-z^k\|^2 - \frac{2}{\gamma_k\theta^2}\mathbb{E}_k[\langle f'(x,B_k), x^k-x\rangle] + \frac{1}{\eta\gamma_k^2\theta^2}\mathbb{E}_k[\|f'(x,B_k)\|^2] + \frac{\beta(1-\beta)}{\theta^2}\|x^k-x^{k-1}\|^2 \\
&\quad - \frac{(1-\beta-\eta)-\tau/\gamma_k}{\theta^2}\mathbb{E}_k[\|x^{k+1}-x^k\|^2].
\end{aligned}
$$

Take $x=x^*$ and by optimality condition we have $\mathbb{E}_k[\langle f'(x^*,B_k), x^k-x^*\rangle] = \langle f'(x^*), x^k-x^*\rangle \le 0$, which implies

$$
\begin{aligned}
\mathbb{E}_k[\|\hat{x}-z^{k+1}\|^2] \le \|\hat{x}-z^k\|^2 + \frac{1}{\eta\gamma_k^2\theta^2}\mathbb{E}_k[\|f'(x^*,B_k)\|^2] + \frac{\beta(1-\beta)}{\theta^2}\|x^k-x^{k-1}\|^2 \\
- \frac{(1-\beta-\eta)-\tau/\gamma_k}{\theta^2}\mathbb{E}_k[\|x^{k+1}-x^k\|^2].
\end{aligned}
$$

Last we take γ_k such that $\beta(1-\beta) \le (1-\beta-\eta)-\tau/\gamma_k \Rightarrow \gamma_k \ge \frac{\tau}{\theta^2-\eta}$ and take summation over $k=0,\dots,K$ to get that

$$
\mathbb{E}[\|\hat{x}-z^{K+1}\|^2] \le \|\hat{x}-z^0\|^2 + \frac{\beta(1-\beta)}{\theta^2}\|x^1-x^0\|^2 + \sum_{k=0}^{K}\frac{1}{\eta\gamma_k^2\theta^2}\mathbb{E}[\|f'(x^*,B)\|^2].
$$

Recall that $\mathbb{E}[\|\hat{x}-z^{K+1}\|^2] = \frac{1}{\theta^2}\|x^*-x^{K+1}\|^2$ and we have

$$
\mathbb{E}[\|x^*-x^{K+1}\|^2] \le \|x^*-x^0\|^2 + \beta(1-\beta)\|x^1-x^0\|^2 + \sum_{k=0}^{K}\frac{1}{\eta\gamma_k^2}\mathbb{E}[\|f'(x^*,B)\|^2].
$$

Our desired result immediately follows by taking $\eta = \theta^2/2$.

References

[Nemirovski et al] Nemirovski, A., et al. “Robust Stochastic Approximation Approach to Stochastic Programming.” Siam Journal on Optimization, vol. 19, no. 4, 2008, pp. 1574–1609.

[AD2019] Asi, Hilal, and John C. Duchi. "Stochastic (approximate) proximal point methods: Convergence, optimality, and adaptivity." *SIAM Journal on Optimization* 29.3 (2019): 2257-2290.