Proof of the Regret Bound of the Proposed OCO

Ning Qi, *Member, IEEE*, Yousuf Baker, *Graduate Student Member, IEEE*, Bolun Xu, *Member, IEEE*

The proposed OCO algorithm is outlined as follows:

$$\min\ f_t(\boldsymbol{x}_t) \quad \text{s.t. } g_t(\boldsymbol{x}_t)\le 0,\ \boldsymbol{x}_t=\{\boldsymbol{x}_t,\boldsymbol{x}_t^\dagger\} \tag{1a}$$

$$Q_{i,t-1}=Q_{i,t-2}+\beta_{i,t-1}[g_{t-1}(\boldsymbol{x}_{i,t-1})]_+ \tag{1b}$$

$$\begin{aligned}\boldsymbol{x}_{i,t}=\operatorname*{argmin}_{\boldsymbol{x}\in\mathcal{X}}\{&\alpha_{i,t-1}\langle\partial f_{t-1}(\boldsymbol{x}_{i,t-1}),\ \boldsymbol{x}-\boldsymbol{x}_{i,t-1}\rangle\\ &+\alpha_{i,t-1}\beta_{t-1}\langle Q_{i,t-1},\ [g_{t-1}(\boldsymbol{x})]_+\rangle+\|\boldsymbol{x}-\boldsymbol{x}_{i,t-1}\|^2\}\end{aligned} \tag{1c}$$

$$\ell_{i,t-1}=\langle\partial f_{t-1}(\boldsymbol{x}_{t-1}),\ \boldsymbol{x}_{i,t-1}-\boldsymbol{x}_{t-1}\rangle \tag{1d}$$

$$\rho_{i,t}=\frac{\rho_{i,t-1}e^{-\gamma\ell_{i,t-1}}}{\sum_{i=1}^{N}\rho_{i,t-1}e^{-\gamma\ell_{i,t-1}}},\ x_t=\sum_{i=1}^{N}\rho_{i,t}\boldsymbol{x}_{i,t} \tag{1e}$$

Theorem 1. Sublinear Dynamic Regret Bound. Given convex functions f_t and g_t defined on a convex, closed set $\mathcal{X}$ with bounded diameter, assume $F,J>0$ exist such that $\forall x,y\in\mathcal{X}$:

$$|f_t(x)-f_t(y)|,\|g_t(x)\|\le F,\quad \|\partial f_t(x)\|,\|\partial g_t(x)\|\le J \tag{2}$$

With parameters set as in (3), we can achieve dynamic regret bound of OCO algorithm in (4).

$$\alpha_{i,t}=\frac{\alpha_0 2^{i-1}}{t^c},\ \beta_{i,t}=\frac{\beta_0}{\sqrt{\alpha_{i,t}}},\ \gamma=\frac{\gamma_0}{T^c},\ N=\lfloor\kappa\log_2(1+T)\rfloor+1 \tag{3}$$

where $\kappa\in[0,c]$, $c\in(0,1)$, $\alpha_0,\beta_0>0$, $\gamma_0\in(0,\frac{1}{\sqrt{2J}})$.

$$\text{Regret}=\mathcal{O}(T^c(1+P_x)^{1-\kappa}+T^{1-c}(1+P_x)^{\kappa}) \tag{4}$$

Proof. □

Let $\{\boldsymbol{x}_{i,t}\}$ and $\{\boldsymbol{x}_t\}$ be the sequences generated by the OCO algorithm. Let $\{\boldsymbol{x}_t^\dagger\}$ be a global optimum in the feasible set $\mathcal{X}$. From f_t is convex and (2), we have:

$$\begin{aligned}&f_t(\boldsymbol{x}_{i,t})-f_t(\boldsymbol{x}_t^\dagger)\le\left\langle\partial f_t(\boldsymbol{x}_{i,t}),\ \boldsymbol{x}_{i,t}-\boldsymbol{x}_t^\dagger\right\rangle\\ &\le G\|\boldsymbol{x}_{i,t}-\boldsymbol{x}_{i,t+1}\|+\left\langle\partial f_t(\boldsymbol{x}_{i,t}),\ \boldsymbol{x}_{i,t+1}-\boldsymbol{x}_t^\dagger\right\rangle\\ &\le\frac{G^2\alpha_{i,t}}{2}+\frac{1}{2\alpha_{i,t}}\|\boldsymbol{x}_{i,t}-\boldsymbol{x}_{i,t+1}\|^2+\left\langle\partial f_t(\boldsymbol{x}_{i,t}),\ \boldsymbol{x}_{i,t+1}-\boldsymbol{x}_t^\dagger\right\rangle\end{aligned} \tag{5}$$

For the rightmost term of (5), we have:

$$\begin{aligned}&\left\langle\partial f_t(\boldsymbol{x}_{i,t}),\ \boldsymbol{x}_{i,t+1}-\boldsymbol{x}_t^\dagger\right\rangle\\ &=\left\langle\beta_{i,t+1}(\partial[g_t(\boldsymbol{x}_{i,t+1})]_+)^TQ_{i,t},\ \boldsymbol{x}_t^\dagger-\boldsymbol{x}_{i,t+1}\right\rangle\\ &+\left\langle\partial f_t(\boldsymbol{x}_{i,t})+\beta_{i,t+1}(\partial[g_t(\boldsymbol{x}_{i,t+1})]_+)^TQ_{i,t},\ \boldsymbol{x}_{i,t+1}-\boldsymbol{x}_t^\dagger\right\rangle\end{aligned} \tag{6}$$

Since g_t is a convex function, it is trivial to show that $[g_t]_+$ is also convex; hence the first term of (6) can be relaxed:

$$\begin{aligned}&\left\langle\beta_{t+1}(\partial[g_t(\boldsymbol{x}_{i,t+1})]_+)^TQ_{i,t},\ \boldsymbol{x}_t^\dagger-\boldsymbol{x}_{i,t+1}\right\rangle\\ &\le\beta_{t+1}\left\langle Q_{i,t},\ [g_t(\boldsymbol{x}_t^\dagger)]_+\right\rangle-\beta_{i,t+1}\langle Q_i(t),\ [g_t(\boldsymbol{x}_{i,t+1})]_+\rangle\end{aligned} \tag{7}$$

From Lemma 1 in [1], we have:

$$\begin{aligned}&\left\langle\partial f_t(\boldsymbol{x}_{i,t})+\beta_{i,t+1}(\partial[g_t(\boldsymbol{x}_{i,t+1})]_+)^TQ_{i,t},\ \boldsymbol{x}_{i,t+1}-\boldsymbol{x}_t^\dagger\right\rangle\\ &\le\frac{1}{\alpha_{i,t}}(\|\boldsymbol{x}_t^\dagger-\boldsymbol{x}_{i,t}\|^2-\|\boldsymbol{x}_t^\dagger-\boldsymbol{x}_{i,t+1}\|^2-\|\boldsymbol{x}_{i,t+1}-\boldsymbol{x}_{i,t}\|^2)\end{aligned} \tag{8}$$

Combining (1e), (5)-(8), we have:

$$\begin{aligned}\ell_t(\boldsymbol{x}_{i,t})-\ell_t(\boldsymbol{x}_t^\dagger)\le&\frac{J^2\alpha_{i,t}}{2}+\frac{1}{\alpha}(\|\boldsymbol{x}_t^\dagger-\boldsymbol{x}_{i,t}\|^2-\|\boldsymbol{x}_t^\dagger-\boldsymbol{x}_{i,t+1}\|^2)\\ &+\beta_{t+1}\left\langle Q_i(t),\ [g_t(\boldsymbol{x}_t^\dagger)]_+\right\rangle\end{aligned} \tag{9}$$

Since the last term of (9) is non-negative, we have:

$$\begin{aligned}&\sum_{t=1}^{T}(\ell_t(\boldsymbol{x}_{i,t})-\ell_t(\boldsymbol{x}_t^\dagger))\le\sum_{t=1}^{T}\frac{J^2\alpha_{i,t}}{2}\\ &+\sum_{t=1}^{T}\frac{1}{\alpha_{i,t}}(\|\boldsymbol{x}_t^\dagger-\boldsymbol{x}_{i,t}\|^2-\|\boldsymbol{x}_t^\dagger-\boldsymbol{x}_{i,t+1}\|^2)\end{aligned} \tag{10}$$

For the first term of (10), we have:

$$\sum_{t=1}^{T}\frac{J^2\alpha_{i,t}}{2}\le\frac{2^{i-1}J^2}{2}\sum_{t=1}^{T}\frac{1}{t^c}\le\frac{2^{i-1}J^2}{2(1-c)}T^{1-c} \tag{11}$$

By leveraging (2) and parameter setting (3), we have:

$$\begin{aligned}&\sum_{t=1}^{T}\frac{t^c}{\alpha_0 2^{i-1}}\Big(\|\boldsymbol{x}_t^\dagger-\boldsymbol{x}_{i,t}\|^2-\|\boldsymbol{x}_t^\dagger-\boldsymbol{x}_{i,t+1}\|^2\Big)\\ &=\frac{1}{\alpha_0 2^{i-1}}\sum_{t=1}^{T}\big(t^c\|\boldsymbol{x}_t^\dagger-\boldsymbol{x}_{i,t}\|^2-(t+1)^c\|\boldsymbol{x}_{t+1}^\dagger-\boldsymbol{x}_{i,t+1}\|^2\\ &+(t+1)^c\|\boldsymbol{x}_{t+1}^\dagger-\boldsymbol{x}_{i,t+1}\|^2-t^c\|\boldsymbol{x}_t^\dagger-\boldsymbol{x}_{i,t+1}\|^2\\ &+t^c\|\boldsymbol{x}_t^\dagger-\boldsymbol{x}_{i,t+1}\|^2-t^c\|\boldsymbol{x}_t^\dagger-\boldsymbol{x}_{i,t}\|^2\big)\\ &\le\frac{1}{\alpha_0 2^{i-1}}\|\boldsymbol{x}_1^\dagger-\boldsymbol{x}_{i,1}\|^2+\frac{1}{\alpha_0 2^{i-1}}\sum_{t=1}^{T}((t+1)^c-t^c)(d(\mathcal{X}))^2\\ &+\frac{2}{\alpha_0 2^{i-1}}\sum_{t=1}^{T}t^c d(\mathcal{X})\|\boldsymbol{x}_{t+1}^\dagger-\boldsymbol{x}_t^\dagger\|\\ &\le\frac{1}{\alpha_0 2^{i-1}}(1+(T+1)^c-1)(d(\mathcal{X}))^2+\frac{2T^c d(\mathcal{X})P_x}{\alpha_0 2^{i-1}}\\ &\le\frac{2}{\alpha_0 2^{i-1}}(d(\mathcal{X}))^2T^c\left(1+\frac{P_x}{d(\mathcal{X})}\right)\end{aligned} \tag{12}$$

Let $i_0=\left\lfloor\frac{1}{2}\log_2(1+\frac{P_x}{d(\mathcal{X})})\right\rfloor+1\in[N]$, such that we have:

$$2^{i_0-1}\le\sqrt{1+\frac{P_x}{d(\mathcal{X})}}\le 2^{i_0}. \tag{13}$$

Combining (11)-(13) yields:

$$\sum_{t=1}^{T}(\ell_t(\boldsymbol{x}_{i_0,t})-\ell_t(\boldsymbol{x}_t^{\dagger}))\leq\frac{4}{\alpha_0}(d(\mathcal{X}))^2T^c\left(1+\frac{P_x}{d(\mathcal{X})}\right)^{1-\kappa}+\frac{J^2\alpha_0}{2(1-c)}T^{1-c}\left(1+\frac{P_x}{d(\mathcal{X})}\right)^{\kappa} \tag{14}$$

Applying Lemma 1 in reference [2] to (1e) yields:

$$\sum_{t=1}^{T}\ell_t(\boldsymbol{x}_t)-\min_{i\in[N]}\{\sum_{t=1}^{T}\ell_t(\boldsymbol{x}_{i,t})+\frac{1}{\gamma}\ln\frac{1}{\rho_{i,1}}\}\leq\frac{\gamma(Jd(\mathcal{X}))^2T}{2} \tag{15}$$

$$\sum_{t=1}^{T}(\ell_t(\boldsymbol{x}_t)-\ell_t(\boldsymbol{x}_{i_0,t}))\leq\frac{\gamma_0(Jd(\mathcal{X}))^2T^{1-c}}{2}+\frac{1}{\gamma_0}T^c\ln\frac{1}{\rho_{i_0,1}} \tag{16}$$

From $\rho_{i,1}=(M+1)/[i(i+1)M]$, we have:

$$\ln\frac{1}{\rho_{i_0,1}}\leq\ln(i_0(i_0+1))\leq 2\ln(i_0+1)\leq 2\ln(\left\lfloor\kappa\log_2(1+\frac{P_x}{d(\mathcal{X})})\right\rfloor) \tag{17}$$

From (1e) and that f_t is convex, we have

$$f_t(\boldsymbol{x}_t)-f_t(\boldsymbol{x}_t^{\dagger})\leq\ell_t(\boldsymbol{x}_t)-\ell_t(\boldsymbol{x}_t^{\dagger}) \tag{18}$$

Combining (14)-(18) yields:

$$\text{Regret}\leq\frac{4}{\alpha_0}(d(\mathcal{X}))^2T^c\left(1+\frac{P_x}{d(\mathcal{X})}\right)^{1-\kappa}+\frac{\gamma_0(Jd(\mathcal{X}))^2T^{1-c}}{2}+\frac{J^2\alpha_0}{2(1-c)}T^{1-c}\left(1+\frac{P_x}{d(\mathcal{X})}\right)^{\kappa}+\frac{2}{\gamma_0}T^c\ln(\lfloor\kappa\log_2\left(1+\frac{P_x}{d(\mathcal{X})}\right)\rfloor) \tag{19}$$

Hence, we finish the proof.

References

[1] X. Yi, X. Li, L. Xie *et al.*, "Distributed online convex optimization with time-varying coupled inequality constraints," *IEEE Transactions on Signal Processing*, vol. 68, pp. 731–746, 2020.

[2] L. Zhang, S. Lu, and Z.-H. Zhou, "Adaptive online learning in dynamic environments," *Advances in neural information processing systems*, vol. 31, 2018.