A. Proofs

A.1. Proof for Theorem 1

Let us apply the given oblivious p-CLI algorithm on a quadratic function of the form

$$f : \mathbb{R}^d \to \mathbb{R} : \ \mathbf{x} \mapsto \frac{1}{2}\mathbf{x}^\top Q\mathbf{x} + \mathbf{q}^\top \mathbf{x},$$

where $Q = \mathrm{diag}\,(\eta, \ldots, \eta)$ and $\mathbf{q} = -\mathbf{v}\eta$ for some $\eta \in [\mu, L]$ and $\mathbf{v} \neq 0 \in \mathbb{R}^d$. In particular, we have that the norm of the unique minimizer is $\|\mathbf{x}^*\| = \left\|-Q^{-1}\mathbf{q}\right\| = \|\mathbf{v}\|$. We set the initialization points to be zero, i.e., $\mathbb{E}\mathbf{x}_j = 0, \ j = 1, \ldots, p$, and denote the corresponding coefficients by $A_{ij}^{(k)}, B_{ij}^{(k)} \in \mathbb{R}^{d\times d}$. The crux of proof is that, as long as η lies in $[\mu, L]$, the side-information $\{\mu, L\}$ is not affected, and therefore, the coefficients remain unchanged.

First, we express $\mathbb{E}\mathbf{x}_i^{k+1}$ in terms of $Q, \mathbf{q}$ and $\mathbb{E}\mathbf{x}_1^{(k)}, \ldots, \mathbb{E}\mathbf{x}_p^{(k)} \in \mathbb{R}^d$. By Definition 3 we have for any $i \in [p]$,

$$\begin{aligned}
\mathbb{E}\mathbf{x}_i^{k+1} &= \sum_{j=1}^{p} \left(A_{ij}^{(k)}\partial f + B_{ij}^{(k)}\right)(\mathbb{E}\mathbf{x}_j^{(k)}) \\
&= \sum_{j=1}^{p}(A_{ij}^{(k)}\partial f(\mathbb{E}\mathbf{x}_j^{(k)}) + B_{ij}^{(k)}\mathbb{E}\mathbf{x}_j^{(k)}) \\
&= \sum_{j=1}^{p}(A_{ij}^{(k)}(Q\mathbb{E}\mathbf{x}_j^{(k)} + \mathbf{q}) + B_{ij}\mathbb{E}\mathbf{x}_j^{(k)}) \\
&= \sum_{j=1}^{p}(A_{ij}^{(k)}Q + B_{ij}^{(k)})\mathbb{E}\mathbf{x}_j^{(k)} + \sum_{j=1}^{p} A_{ij}^{(k)}\mathbf{q}.
\end{aligned}$$

Our next step is to reduce the problem of minimizing f to a polynomial approximation problem. We claim that for any $k \geq 1$ and $i \in [d]$ there exist d real polynomials $s_{k,i,1}(\eta), \ldots, s_{k,i,d}(\eta)$ of degree at most $k-1$, such that

$$\mathbb{E}\mathbf{x}_i^{(k)} = (s_{k,i,*}(\eta))\eta, \tag{17}$$

where

$$(s_{k,i,*}(\eta)) \coloneqq (s_{k,i,1}(\eta), \ldots, s_{k,i,d}(\eta))^\top.$$

Let us prove this claim using mathematical induction. For $k = 1$ we have

$$\mathbb{E}\mathbf{x}_i^{(1)} = \sum_{j=1}^{p}(A_{ij}^{(0)}Q + B_{ij}^{(0)})\mathbb{E}\mathbf{x}_j^{(0)} + \sum_{j=1}^{p} A_{ij}^{(0)}\mathbf{q} = -\sum_{j=1}^{p} A_{ij}^{(0)}\mathbf{v}\eta, \tag{18}$$

showing that the base case holds. For the induction step, assume the statement holds for some $k > 1$ with $s_{k,i,j}(\eta)$ as above, then

$$\begin{aligned}
\mathbb{E}\mathbf{x}_i^{(k+1)} &= \sum_{j=1}^{p}(A_{ij}^{(k)}Q + B_{ij}^{(k)})\mathbb{E}\mathbf{x}_j^{(k)} + \sum_{j=1}^{p} A_{ij}^{(k)}\mathbf{q} \\
&= \sum_{j=1}^{p}(A_{ij}^{(k)}\mathrm{diag}\,(\eta, \ldots, \eta) + B_{ij}^{(k)})(s_{k,j,*}(\eta))\eta - \sum_{j=1}^{p} A_{ij}^{(k)}\mathbf{v}\eta \\
&= \left(\sum_{j=1}^{p}(A_{ij}^{(k)}\mathrm{diag}\,(\eta, \ldots, \eta) + B_{ij}^{(k)})(s_{k,j,*}(\eta)) - \sum_{j=1}^{p} A_{ij}^{(k)}\mathbf{v}\right)\eta.
\end{aligned} \tag{19}$$

The expression inside the last parenthesis is a vector d entries, each of which contains a real polynomial of degree at most k. This concludes the induction step (note that the derivations of equalities (18) and (19) above are exactly where we use the fact that there is no functional dependency of $A_{ij}^{(k)}$ and $B_{ij}^{(k)}$ on η).

We are now ready to estimate the sub-optimality of $\mathbb{E}\mathbf{x}_p^{(k)}$, the point returned by the algorithm at the k'th iteration. Setting $m \in \operatorname{argmax}_j |\mathbf{v}_j|$ we have

$$\left\|\mathbb{E}\mathbf{x}_p^{(k)} - \mathbf{x}^*\right\| = \|(s_{k,p,*}(\eta))\eta + \mathbf{v}\| \geq |s_{k,p,m}(\eta)\eta + v_m| = |v_m||s_{k,p,m}(\eta)\eta/v_m + 1|. \tag{20}$$

By Lemma 2 in appendix B, there exists $\eta \in [\mu, L]$, such that

$$|s_{k,p,m}(\eta)\eta/v_m + 1| \geq \left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^k,$$

where $\kappa = L/\mu$. Defining Q and $\mathbf{q}$ accordingly, and choosing, e.g., $\mathbf{v} = R\mathbf{e}_1$ where R denotes a prescribed distance, yields

$$\left\|\mathbb{E}\mathbf{x}_p^{(k)} - \mathbf{x}^*\right\| \geq R\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^k.$$

Lastly, using the fact that f is μ-strongly convex concludes the proof of the first part of the theorem.

For the smooth case we need to estimate $f(\mathbb{E}\mathbf{x}_p^{(k)}) - f^*$. Let $\eta \in (0, L]$ and define Q and $\mathbf{q}$, accordingly. Inequality (20) yields

$$\begin{aligned} f(\mathbb{E}\mathbf{x}^{(k)}) - f(\mathbf{x}^*) &= \frac{1}{2}(\mathbb{E}\mathbf{x}^{(k)} - \mathbf{x}^*)^\top Q(\mathbb{E}\mathbf{x}^{(k)} - \mathbf{x}^*) \\ &\geq \frac{(v_m)^2}{2}\eta((s_{k,p,m})(\eta)\eta/v_m + 1)^2. \end{aligned} \tag{21}$$

Now, by Lemma 3 in appendix B,

$$\min_{s(\eta),\ \partial s \leq k-1} \max_{\eta \in (0,L]} \eta(s(\eta)\eta + 1)^2 \geq \frac{L}{(2k+1)^2}$$

Thus, choosing $\mathbf{v} = R\mathbf{e}_1$ concludes the proof.

A.2. Proof for Theorem 2

The proof of this theorem follows the exact reduction used in the proof of Theorem 1 (see Appendix A.1 above). The only difference is that here μ is allowed to be any real number in $(0, L)$. This consideration reduces our problem into, yet another, polynomial approximation problem. For completeness, we provide here the full proof.

Let us apply the given oblivious p-CLI algorithm on a quadratic function of the form

$$f : \mathbb{R}^d \to \mathbb{R} :\ \mathbf{x} \mapsto \frac{1}{2}\mathbf{x}^\top Q\mathbf{x} + \mathbf{q}^\top \mathbf{x},$$

where $Q = \operatorname{diag}(\eta, \ldots, \eta)$ and $\mathbf{q} = -\mathbf{v}\eta$ for some $\eta \in (0, L)$ and $\mathbf{v} \neq 0 \in \mathbb{R}^d$. In particular, we have that the norm of the unique minimizer is $\|\mathbf{x}^*\| = \left\|-Q^{-1}\mathbf{q}\right\| = \|\mathbf{v}\|$. We set the initialization points to be zero, i.e., $\mathbb{E}\mathbf{x}_j = 0,\ j = 1, \ldots, p$, and denote the corresponding coefficients by $A_{ij}^{(k)}, B_{ij}^{(k)} \in \mathbb{R}^{d\times d}$. The crux of proof is that, as long as η lies in $(0, L]$, the side-information $\{\mu, L\}$ is not affected, and therefore, the coefficients remain unchanged.

First, we express $\mathbb{E}\mathbf{x}_i^{k+1}$ in terms of $Q, \mathbf{q}$ and $\mathbb{E}\mathbf{x}_1^{(k)}, \ldots, \mathbb{E}\mathbf{x}_p^{(k)} \in \mathbb{R}^d$. By Definition 3 we have for any $i \in [p]$,

$$\begin{aligned}\mathbb{E}\mathbf{x}_i^{k+1} &= \sum_{j=1}^{p} \left(A_{ij}^{(k)} \partial f + B_{ij}^{(k)}\right)(\mathbb{E}\mathbf{x}_j^{(k)}) \\ &= \sum_{j=1}^{p} (A_{ij}^{(k)} \partial f(\mathbb{E}\mathbf{x}_j^{(k)}) + B_{ij}^{(k)} \mathbb{E}\mathbf{x}_j^{(k)}) \\ &= \sum_{j=1}^{p} (A_{ij}^{(k)} (Q\mathbb{E}\mathbf{x}_j^{(k)} + \mathbf{q}) + B_{ij} \mathbb{E}\mathbf{x}_j^{(k)}) \\ &= \sum_{j=1}^{p} (A_{ij}^{(k)} Q + B_{ij}^{(k)}) \mathbb{E}\mathbf{x}_j^{(k)} + \sum_{j=1}^{p} A_{ij}^{(k)} \mathbf{q}\end{aligned}$$

Our next step is to reduce the problem of minimizing f to a polynomial approximation problem. We claim that for any $k \geq 1$ and $i \in [d]$ there exist d real polynomials $s_{k,i,1}(\eta), \ldots, s_{k,i,d}(\eta)$ of degree at most $k-1$, such that

$$\mathbb{E}\mathbf{x}_i^{(k)} = (s_{k,i,*}(\eta))\eta, \tag{22}$$

where

$$(s_{k,i,*}(\eta)) \coloneqq (s_{k,i,1}(\eta), \ldots, s_{k,i,d}(\eta))^\top.$$

Let us prove this claim using mathematical induction. For $k = 1$ we have,

$$\mathbb{E}\mathbf{x}_i^{(1)} = \sum_{j=1}^{p} (A_{ij}^{(0)} Q + B_{ij}^{(0)}) \mathbb{E}\mathbf{x}_j^{(0)} + \sum_{j=1}^{p} A_{ij}^{(0)} \mathbf{q} = -\sum_{j=1}^{p} A_{ij}^{(0)} \mathbf{v}\eta, \tag{23}$$

showing that the base case holds. For the induction step, assume the statement holds for some $k > 1$ with $s_{k,i,j}(\eta)$ as above, then

$$\begin{aligned}\mathbb{E}\mathbf{x}_i^{(k+1)} &= \sum_{j=1}^{p} (A_{ij}^{(k)} Q + B_{ij}^{(k)}) \mathbb{E}\mathbf{x}_j^{(k)} + \sum_{j=1}^{p} A_{ij}^{(k)} \mathbf{q} \\ &= \sum_{j=1}^{p} (A_{ij}^{(k)} \mathrm{diag}\,(\eta, \ldots, \eta) + B_{ij}^{(k)})(s_{k,j,*}(\eta))\eta - \sum_{j=1}^{p} A_{ij}^{(k)} \mathbf{v}\eta \\ &= \left(\sum_{j=1}^{p} (A_{ij}^{(k)} \mathrm{diag}\,(\eta, \ldots, \eta) + B_{ij}^{(k)})(s_{k,j,*}(\eta)) - \sum_{j=1}^{p} A_{ij}^{(k)} \mathbf{v} \right) \eta. \end{aligned} \tag{24}$$

The expression inside the last parenthesis is a vector d entries, each of which contains a real polynomial of degree at most k. This concludes the induction step. (note that the derivations of equalities (23) and (24) above are exactly where we use the fact that there is no functional dependency of $A_{ij}^{(k)}$ and $B_{ij}^{(k)}$ on η).

We are now ready to estimate the sub-optimality of $\mathbb{E}\mathbf{x}_p^{(k)}$, the point returned by the algorithm at the k'th iteration. Let us set $m \in \mathrm{argmax}_j |\mathbf{v}_j|$, then

$$\begin{aligned}\left\|\mathbb{E}\mathbf{x}_p^{(k)} - \mathbf{x}^*\right\| &= \|(s_{k,p,*}(\eta))\eta + \mathbf{v}\| \\ &\geq |s_{k,p,m}(\eta)\eta + v_m| \\ &= |v_m||s_{k,p,m}(\eta)\eta/v_m + 1|. \end{aligned} \tag{25}$$

By Lemma 4, there exists $\eta \in (L/2, L)$, such that

$$|s_{k,p,m}(\eta)\eta/v_m + 1| \geq (1 - \eta/L)^{k+1}. \tag{26}$$

Defining Q and $\mathbf{q}$ accordingly, and choosing, e.g., $\mathbf{v} = R\mathbf{e}_1$ where R denotes a prescribed distance, yields

$$\left\|\mathbb{E}\mathbf{x}_p^{(k)} - \mathbf{x}^*\right\| \geq (1 - \eta/L)^{k+1}.$$

Using the fact that f is $L/2$-strongly convex concludes the proof.

B. Technical Lemmas

Below, we provide 3 lemmas which are used to bound from below the quantity $|s(\eta)\eta+1|$ over different domains of η, where $s(\eta)$ is a real polynomial. For brevity, we denote the set of real polynomials of degree k by $\mathcal{P}_k$.

Lemma 2. *Let $s(\eta)\in\mathcal{P}_k$, and let $0<\mu<L$. Then,*

$$\max_{\eta\in[\mu,L]}|s(\eta)\eta+1|\geq\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^{k+1}$$

where $\kappa:=L/\mu$.

Proof Denote $q(\eta):=T_{k+1}^{-1}\left(\frac{L+\mu}{L-\mu}\right)T_{k+1}\left(\frac{2\eta-\mu-L}{L-\mu}\right)$, where $T_k(\eta)$ denotes the Chebyshev polynomial of degree k,

$$T_k(\eta)=\begin{cases}\cos(k\arccos(\eta)) & |\eta|\leq 1\\ \cosh(k\,\text{arcosh}(\eta)) & \eta\geq 1\\ (-1)^n\cosh(k\,\text{arcosh}(-\eta)) & \eta\leq -1.\end{cases}\tag{27}$$

It follows that $|T_{k+1}(\eta)|\leq 1$ for $\eta\in[-1,1]$ and

$$T_{k+1}(\cos(j\pi/(k+1)))=(-1)^j,\ j=0,\dots,k+1.$$

Accordingly, $|q(\eta)|\leq T_{k+1}^{-1}\left(\frac{L+\mu}{L-\mu}\right),\ \eta\in[\mu,L]$ and

$$q\left(\theta_j\right)=(-1)^jT_{k+1}^{-1}\left(\frac{L+\mu}{L-\mu}\right),\ j=0,\dots,k+1,$$

where

$$\theta_j=\frac{\cos(j\pi/(k+1))(L-\mu)+\mu+L}{2}.$$

Suppose, for the sake of contradiction, that

$$\max_{\eta\in[\mu,L]}|s(\eta)\eta+1|<\max_{\eta\in[\mu,L]}|q(\eta)|.$$

Thus, for $r(\eta)=q(\eta)-(1+s(\eta)\eta))$, we have $r(\theta_j)>0$ for even j, and $r(\theta_j)<0$ for odd j. Hence, $r(\eta)$ has $k+1$ roots in $[\mu,L]$. But, since $r(0)=0$ and $\mu>0$, it follows $r(\eta)$ has at least $k+2$ roots, which contradicts the fact that the degree of $r(\eta)$ is at most $k+1$. Therefore,

$$\max_{\eta\in[\mu,L]}|s(\eta)\eta+1|\geq\max_{\eta\in[\mu,L]}|q(\eta)|=T_{k+1}^{-1}\left(\frac{\kappa+1}{\kappa-1}\right),$$

where $\kappa=L/\mu$. Since $(\kappa+1)/(\kappa-1)\geq 1$, we have by Equation (27),

$$\begin{aligned}T_k\left(\frac{\kappa+1}{\kappa-1}\right)&=\cosh\left(k\,\text{arcosh}\left(\frac{\kappa+1}{\kappa-1}\right)\right)\\&=\cosh\left(k\ln\left(\frac{\kappa+1}{\kappa-1}+\sqrt{\left(\frac{\kappa+1}{\kappa-1}\right)^2-1}\right)\right)\\&=\cosh\left(k\ln\left(\frac{\kappa+2\sqrt{\kappa}+1}{\kappa-1}\right)\right)\\&=\cosh\left(k\ln\left(\frac{\sqrt{\kappa}+1}{\sqrt{\kappa}-1}\right)\right)\\&=\frac{1}{2}\left(\left(\frac{\sqrt{\kappa}+1}{\sqrt{\kappa}-1}\right)^k+\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^k\right)\\&\leq\left(\frac{\sqrt{\kappa}+1}{\sqrt{\kappa}-1}\right)^k.\end{aligned}$$

Hence,

$$\max_{\eta\in[\mu,L]} |s(\eta)\eta+1| \geq \left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^{k+1}$$

■

Lemma 3. *Let $s(\eta)\in\mathcal{P}_k$, and let $0<L$. Then,*

$$\max_{\eta\in[0,L]} \eta|s(\eta)\eta+1|^2 \geq \frac{L}{(2k+3)^2}$$

Proof First, we define

$$q(\eta) = \begin{cases} (-1)^k(2k+3)^{-1}\sqrt{L/\eta}T_{2k+3}(\sqrt{\eta/L}) & \eta\neq 0\\ 0 & \eta = 0\end{cases}$$

where $T_k(\eta)$ is the k'th Chebyshev polynomial (see (27)). Let us show that $q(\eta)$ is a polynomial of degree $k+1$ and that $q(0)=1$. The following trigonometric identity

$$\cos\alpha+\cos\beta = 2\cos\left(\frac{\alpha-\beta}{2}\right)\cos\left(\frac{\alpha+\beta}{2}\right)$$

together with (27), yields the following recurrence formula

$$T_k(\eta) = 2\eta T_{k-1}(\eta) - T_{k-2}(\eta).$$

Noticing that $T_0(\eta)=1$ and $T_1(\eta)=x$ (also by (27)), we can use mathematical induction to prove that Chebyshev polynomials of odd degree have only odd powers and that the corresponding coefficient for the first power η in $T_{2k+3}(\eta)$ is indeed $(-1)^k(2k+3)$. Equivalently, we get that $q(\eta)$ is a polynomial of degree $k+1$ and that $q(0)=1$. Next, note that for

$$\theta_j = L\cos\left(\frac{j\pi}{2k+3}\right)^2 \in [0,L], \quad j=0,\ldots,k+1$$

we have

$$\max_{\eta\in[0,L]} \eta^{1/2}|q(\eta)| = (-1)^j\theta_j^{1/2}q(\theta_j) = \frac{\sqrt{L}}{2k+3}.$$

Now, suppose, for the sake of contradiction, that

$$\max_{\eta\in[0,L]} \eta|s(\eta)\eta+1|^2 < \max_{\eta\in[0,L]} \eta|q(\eta)|^2.$$

In particular,

$$\theta_j^{1/2}|s(\theta_j)\theta_j^{1/2}+1| < \theta_j^{1/2}|q(\theta_j)|.$$

Since $\theta_j>0$, we have

$$|s(\theta_j)\theta_j^{1/2}+1| < |q(\theta_j)|.$$

We proceed in a similar way to the proof of Lemma 2. For $r(\eta)=q(\eta)-(1+s(\eta)\eta))$, we have $r(\theta_j)>0$ for even j, and $r(\theta_j)<0$ for odd j. Hence, $r(\eta)$ has $k+1$ roots in $[\theta_{k+1},L]$. But, since $r(0)=0$ and $\theta_{k+1}>0$, it follows $r(\eta)$ has at least $k+2$ roots, which contradicts the fact that degree of $r(\eta)$ is a at most $k+1$.

Therefore,

$$\max_{\eta\in[0,L]} \eta|s(\eta)\eta+1|^2 \geq \max_{\eta\in[0,L]} \eta|q(\eta)|^2 \geq \frac{L}{(2k+3)^2}$$

which concludes the proof.

■

Lemma 4. *Let $s(\eta)\in\mathcal{P}_k$, and let $0<L$. Then exactly one of the two following holds:*

1. *For any $\epsilon>0$, there exists $\eta\in(L-\epsilon,L)$ such that*

$$|s(\eta)\eta+1| > (1-\eta/L)^{k+1}.$$

2. $s(\eta)\eta+1=(1-\eta/L)^{k+1}$.

Proof It suffices to show that if (1) does not hold then $s(\eta)\eta+1=(1-\eta/L)^{k+1}$. Suppose that there exists $\epsilon>0$ such that for all $\eta\in(L-\epsilon,L)$ it holds that

$$|s(\eta)\eta+1| \leq \left(1-\frac{\eta}{L}\right)^{k+1}.$$

Define

$$q(\eta) \coloneqq s\left(L(1-\eta)\right)L(1-\eta)+1 \tag{28}$$

and denote the corresponding coefficients by $q(\eta)=\sum_{j=0}^{k+1} q_i\eta^j$. We show by induction that $q_j=0$ for all $j=0,\dots,k$. For $j=0$ we have that since for any $\eta\in(0,1-(L-\epsilon)/L)$

$$|q(\eta)| \leq \left(1-\frac{L(1-\eta)}{L}\right)^{k+1} = \eta^{k+1},$$

it holds that

$$|q_0| = |q(0)| = \left|\lim_{\eta\to 0^+} q(\eta)\right| \leq \lim_{\eta\to 0^+} \eta^{k+1} = 0.$$

Now, if $q_0=\dots=q_{m-1}=0$ for $m<k+1$ then

$$|q_m| = \left|\frac{q(0)}{\eta^m}\right| = \left|\lim_{\eta\to 0^+}\frac{q(\eta)}{\eta^m}\right| \leq \lim_{t\to 0^+} \eta^{k+1-m} = 0.$$

Thus, proving the induction claim. This, in turns, implies that $q(\eta)=q_{k+1}\eta^{k+1}$. Now, by Equation (28), it follows that $q_{k+1}=q(1)=1$. Hence, $q(\eta)=\eta^{k+1}$. Lastly, using Equation (28) again yields

$$s(\eta)\eta+1 = q\left(1-\frac{\eta}{L}\right) = \left(1-\frac{\eta}{L}\right)^{k+1},$$

concluding the proof. ■