Convergence rate

Note that we have:

$H(x,p) = k(p) + f(x) - f(x_*)$

$\mathcal{V}(x,p) = H(x,p) + \beta \langle x - x_*, p \rangle$

f: convex, unique minimum at x_*

k: strictly convex, minimum $k(0) = 0$, $k(p_t) \geq \alpha f_c^*(-p_t)$

$r \in (0,1)$, $\alpha \in (0,1]$, $\beta \in (0, \min(\alpha, \gamma)]$, $\lambda \in (0, \infty)$

Proof:

$\mathcal{V}_t' = -\gamma \langle \nabla k(p_t), p_t \rangle + \beta \langle \nabla k(p_t), p_t \rangle - \beta r \langle x_t - x_*, p_t \rangle - \beta \langle x_t - x_*, \nabla f(x_t) \rangle$

$= -(\gamma - \beta) \langle \nabla k(p_t), p_t \rangle - \beta \gamma \langle x_t - x_*, p_t \rangle - \beta \langle x_t - x_*, \nabla f(x_t) \rangle$

$\because$ $k(p_t) \leq \langle \nabla k(p_t) p_t \rangle$ and $f(x_t) - f(x_*) \leq \langle x_t - x_*, \nabla f(x_t) \rangle$

$\therefore$ $\mathcal{V}_t' \leq -(r - \beta) k(p_t) - \beta r \langle x_t - x_*, p_t \rangle - \beta (f(x_t) - f(x_*))$

$\because$ convexity of $\mathcal{V}_t$ and $\beta \leq r$

$\therefore$ Our goal is to show that $\mathcal{V}_t' = -\lambda \mathcal{V}_t$ for some rate $\lambda > 0$

$\mathcal{V}_t' \leq -\lambda \mathcal{V}_t$

$-(\gamma - \beta) k(p_t) - \beta r \langle x_t - x_*, p_t \rangle - \beta (f(x_t) - f(x_*)) \leq -\lambda (k(p_t) + f(x_t) - f(x_*) + \beta \langle x_t - x_*, p_t \rangle)$

rearrange it, we have:

$-\beta (r - \lambda) \langle x_t - x_*, p_t \rangle \leq (\gamma - \beta - \lambda) k(p_t) + (\beta - \lambda)(f(x_t) - f(x_*))$ ①

$\boxed{\text{Assume that } \lambda \leq r}$ (as long as we can find such λ, we are good)

Based on (15) from paper: $\langle x - x_*, p \rangle \geq -(k(p)/\alpha + f(x) - f(x_*))$, we have:

$k(p_t) \geq -\alpha \langle x_t - x_*, p_t \rangle - \alpha (f(x_t) - f(x_*))$

$\therefore$ $k(p_t) + \alpha (f(x_t) - f(x_*)) \geq -\alpha \langle x_t - x_*, p_t \rangle$

Multiply $\frac{\beta}{\alpha}(r - \lambda)$ on both sides $\left(\frac{\beta}{\alpha}(r - \lambda) \geq 0\right)$:

$-\beta (\gamma - \lambda) \langle x_t - x_*, p_t \rangle \leq \frac{\beta}{\alpha}(r - \lambda)\left(k(p_t) + \alpha (f(x_t) - f(x_*))\right)$ ②

$\because$ $k(p_t) \geq 0$ and $f(x_t) - f(x_*) \geq 0$

$\therefore$ compare ① and ②, we have:

$$\begin{cases} \frac{\beta}{\alpha}(r - \lambda) \leq \gamma - \beta - \lambda \\ \beta (r - \lambda) \leq \beta - \lambda \end{cases} \Rightarrow \begin{cases} \lambda \leq \frac{\alpha r - \alpha\beta - \beta r}{\alpha - \beta} \\ \lambda \leq \frac{\beta(1-r)}{1-\beta} \\ \lambda \leq r \text{ (from assumption)} \end{cases} \Rightarrow \lambda \leq \min\left(r, \frac{\alpha r - \alpha\beta - \beta r}{\alpha - \beta}, \frac{\beta(1-r)}{1-\beta}\right)$$

$\because$ $0 < \beta \leq r < 1$

$\therefore$ $\frac{\beta}{1-\beta}(1 - r) \leq r$

$\therefore$ $\lambda(\alpha, \beta, r) = \min\left(\frac{\alpha r - \alpha\beta - \beta r}{\alpha - \beta}, \frac{\beta(1-r)}{1-\beta}\right)$

$\because$ we want optimal rate λ, the bigger λ, the better.

$\therefore$ we need to maximize $\lambda(\alpha, \beta, r)$ in β.

We can add another constraint on β : $\boxed{0 < \beta < \frac{\alpha r}{\alpha + r}}$

So in this interval, we notice $\frac{\alpha r - \alpha\beta - \beta r}{\alpha - \beta}$ is strictly $\searrow$, $\frac{\beta(1-r)}{1-\beta}$ is strictly $\nearrow$.

$\therefore$ the λ_{max} is when two terms are equal.

$\therefore \beta_{\pm} = \frac{1}{1+\alpha}\left(\alpha + \frac{r}{2} \pm \sqrt{(1-r)\alpha^2 + \frac{r^2}{4}}\right)$

$\because$ one can check that $\beta_+ > \frac{\alpha r}{\alpha + r}$ while $0 < \beta_- < \frac{\alpha r}{\alpha + r}$.

$\therefore \max\limits_{\beta \in (0,\alpha]} \lambda(\alpha, \beta, r) = \lambda(\alpha, \beta_-, r)$ for $\alpha \leq 1$

For $\alpha = 1$, we have $\beta^* = \beta_- = \frac{r}{2}$, and $\lambda^* = \frac{r(1-r)}{2-r}$

$\boxed{\text{Assume } \beta \in (0, \frac{\alpha r}{2}]}$

We have proved that $\lambda(\alpha, r, \beta) = \frac{\beta(1-r)}{1-\beta}$ for $\beta < \beta_-$

We just have to prove $\boxed{\beta_- > \frac{\alpha r}{2}}$ to get our result

$\because \beta_-$ can be viewed as a function of r.

$\therefore \beta_-(r)$ is strictly concave with $\beta_-(0) = 0$, $\beta(1) = \frac{\alpha}{1+\alpha}$

$\therefore \boxed{\beta_-} = \beta_-(r) > r\beta_-(1) = \frac{r\alpha}{1+\alpha} \geq \boxed{\frac{\alpha r}{2}} \geq \beta$

Based on ①, we know that:

$$-\beta(r-\lambda)\langle x_t - x_*, p_t\rangle \leq (r - \beta - r^2(1-r)/4 - \lambda)k(p_t) + (\beta - \lambda)(f(x_t) - f(x_*))$$

Compare with ②, we have:

$$\begin{cases} \frac{\beta}{\alpha}(r-\lambda) \leq r - r^2(1-r)/4 - \beta - \lambda \\ \beta(r-\lambda) \leq \beta - \lambda \text{ (already verified)} \end{cases} \Rightarrow r - \beta - \lambda - \frac{\beta}{\alpha}(r-\lambda) \geq r^2(1-r)/4 \text{ for every } \begin{cases} 0 < r < 1 \\ 0 < \alpha \leq 1 \\ 0 < \beta \leq \frac{\alpha r}{2} \end{cases}$$

It is easy to see that we only need to check this for $\beta = \frac{\alpha r}{2}$, and in this case by minimizing the left hand side for $0 < \alpha \leq 1$ and using $\lambda = (1-r)\beta$, we can get the result.