ISyE8813
Convergence of the HPE

Renato Monteiro

March 6, 2024

1 Convergence of the HPE framework

Recall the HPE framework

Inexact IPP

0) Let $z_0 \in \mathbb{R}^n$ and $\sigma \in [0, 1]$ be given and set $k = 1$.

1) Choose $\lambda_k > 0$ and find $(x_k, \tilde{x}_k, \varepsilon_k) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}_+$ such that

$$\tilde{v}_k := \frac{z_{k-1} - z_k}{\lambda_k} \in T^{\varepsilon_k}(\tilde{z}_k)$$
$$\|\tilde{z}_k - z_k\|^2 + 2\lambda_k\varepsilon_k \leq \sigma^2 \|\tilde{z}_k - z_{k-1}\|^2$$

2) Set $k \leftarrow k + 1$ and go to step 1.

Pointwise Convergence

Lemma 1: For every $k \geq 1$ and $z \in \mathbb{R}^n$, have

$$\begin{aligned}&\|z - z_{k-1}\|^2 - \|z - z_k\|^2 \\ &= 2\lambda_k \langle \tilde{z}_k - z, \tilde{v}_k \rangle + \|\tilde{z}_k - z_{k-1}\|^2 - \|\tilde{z}_k - z_k\|^2\end{aligned}$$

Proof: Have

$$\begin{aligned}&\|z - z_{k-1}\|^2 - \|z - z_k\|^2 \\ &= 2\langle z - z_k, z_k - z_{k-1} \rangle + \|z_k - z_{k-1}\|^2 \\ &= 2\langle z - \tilde{z}_k, z_k - z_{k-1} \rangle + 2\langle \tilde{z}_k - z_k, z_k - z_{k-1} \rangle + \|z_k - z_{k-1}\|^2 \\ &= 2\langle z - \tilde{z}_k, z_k - z_{k-1} \rangle + \|\tilde{z}_k - z_{k-1}\|^2 - \|\tilde{z}_k - z_k\|^2 \\ &= -2\lambda_k \langle z - \tilde{z}_k, \tilde{v}_k \rangle + \|\tilde{z}_k - z_{k-1}\|^2 - \|\tilde{z}_k - z_k\|^2\end{aligned}$$

Lemma 2: For every $k \geq 1$ and $z_* \in T^{-1}(0)$, have

$$\|z_* - z_{k-1}\|^2 - \|z_* - z_k\|^2 \geq (1 - \sigma^2)\|\tilde{z}_k - z_{k-1}\|^2 \geq 0$$

Proof: If $z_* \in T^{-1}(0)$ then it follows that $(z_*, 0) \in \operatorname{gr} T$. Moreover, since $\tilde{v}_k \in T^{\varepsilon_k}(\tilde{z}_k)$, we have $(\tilde{z}_k, \tilde{v}_k) \in \operatorname{gr} T^{\varepsilon_k}$. Hence, it follows from the definition of T^{ε_k} that

$$\langle \tilde{z}_k - z_*, \tilde{v}_k \rangle = \langle \tilde{z}_k - z_*, \tilde{v}_k - 0 \rangle \geq -\varepsilon_k$$

This inequality and Lemma 1 with $z = z_*$ then imply that

$$\begin{aligned}&\|z_* - z_{k-1}\|^2 - \|z_* - z_k\|^2 \\ &= 2\lambda_k \langle \tilde{z}_k - z_*, \tilde{v}_k \rangle + \|\tilde{z}_k - z_{k-1}\|^2 - \|\tilde{z}_k - z_k\|^2 \\ &\geq -2\lambda_k \varepsilon_k + \|\tilde{z}_k - z_{k-1}\|^2 - \|\tilde{z}_k - z_k\|^2 \\ &= \|\tilde{z}_k - z_{k-1}\|^2 - \left[\|\tilde{z}_k - z_k\|^2 + 2\lambda_k \varepsilon_k\right] \\ &\geq \|\tilde{z}_k - z_{k-1}\|^2 - \sigma^2 \|\tilde{z}_k - z_{k-1}\|^2\end{aligned}$$

where the last inequality is due to the HPE inequality condition

Lemma 3: For every $K \geq 1$ and $z_* \in T^{-1}(0)$, have

$$\|z_* - z_0\|^2 - \|z_* - z_K\|^2 \geq (1 - \sigma^2) \sum_{k=1}^{K} \|\tilde{z}_k - z_{k-1}\|^2$$

Proof: Follows by adding the inequality in Lemma 2 from $k = 1$ to $k = K$.

Lemma 4: For every $K \geq 1$, have

$$\max \left\{ \frac{2\lambda_k \varepsilon_k}{\sigma^2}, \frac{\lambda_k^2 \|\tilde{v}_k\|^2}{(1+\sigma)^2} \right\} \leq \|\tilde{z}_k - z_{k-1}\|^2$$

Proof: The first inequality of the Lemma follows from the HPE inequality condition. The second one is due to the fact that

$$\begin{aligned}
\lambda_k \|\tilde{v}_k\| &= \|z_k - z_{k-1}\| \\
&= \|z_k - \tilde{z}_k + \tilde{z}_k - z_{k-1}\| \\
&\leq \|z_k - \tilde{z}_k\| + \|\tilde{z}_k - z_{k-1}\| \\
&\leq \sigma \|\tilde{z}_k - z_{k-1}\| + \|\tilde{z}_k - z_{k-1}\| \\
&= (1+\sigma)\|\tilde{z}_k - z_{k-1}\|
\end{aligned}$$

where the second inequality is due to the HPE inequality condition

Lemma 5: For every $K \geq 1$, there exists $k \in \{1, \dots, K\}$ such that

$$\varepsilon_k \leq \frac{\sigma^2 d_0^2}{2\lambda_k(1-\sigma^2)K},$$
$$\|\tilde{v}_k\|^2 \leq \frac{(1+\sigma)^2 d_0^2}{\lambda_k^2(1-\sigma^2)K}$$

where $d_0 = \min\{\|z_* - z_0\| : z_* \in T^{-1}(0)\}$.

As a consequence, if there exists $\underline{\lambda} > 0$ such that $\lambda_k \geq \underline{\lambda}$ for every $k \in \{1, \dots, K\}$, then there exists $k \in \{1, \dots, K\}$ such that

$$\max\left\{\varepsilon_k, \|\tilde{v}_k\|^2\right\} = \mathcal{O}\left(\frac{1}{K}\right)$$

Proof: Let z_* be the closest point to z_0 lying in $T^{-1}(0)$. Hence, $d_0 = \|z_* - z_0\|$. Then, it follows from Lemma 3 that

$$K \min_{k=1,\dots,K} \|\tilde{z}_k - z_{k-1}\|^2 \leq \sum_{k=1}^{K} \|\tilde{z}_k - z_{k-1}\|^2 \leq \frac{d_0^2}{1-\sigma^2}$$

Thus, there exists $k \in \{1, \dots, K\}$ such that

$$\frac{d_0^2}{K(1-\sigma^2)} \geq \|\tilde{z}_k - z_{k-1}\|^2 \geq \max\left\{\frac{2\lambda_k \varepsilon_k}{\sigma^2}, \frac{\lambda_k^2 \|\tilde{v}_k\|^2}{(1+\sigma)^2}\right\}$$

where the last inequality is due to Lemma 4.

Ergodic Convergence

Technical Lemma: Assume that T is maximal monotone and that $\operatorname{gr} T^\varepsilon \neq \emptyset$ for some $\varepsilon \in \mathbb{R}$. Then, $\varepsilon \geq 0$

Proof: Assume for contradiction that $(y, w) \in \operatorname{gr} T^\varepsilon$ for some $\varepsilon < 0$ and $(y, w) \in \mathbb{R}^n \times \mathbb{R}^n$. Then

$$\langle w - v, y - x \rangle \geq -\varepsilon > 0 \quad \forall (x, v) \in \operatorname{gr} T$$

Have $(w, y) \in \operatorname{gr} T$ since otherwise the above inequality with $(x, v) = (y, w)$ would give an absurd. Consider now the multi-valued operator $\tilde{T}$ such that

$$\operatorname{gr} \tilde{T} = T \cup \{(w, y)\}$$

It follows from the monotonicity of T and the above inequality that $\tilde{T}$ is monotone. Moreover, $\tilde{T}$ is larger than T and includes T. These two conclusions then contradict the assumption that T is a maximal monotone operator

Lemma 6: For every $K \geq 1$, we have

$$\tilde{v}_K^a \in T^{\varepsilon_K^a}(\tilde{z}_K^a) \tag{1}$$

where

$$\tilde{v}_K^a := \frac{1}{\Lambda_K}\sum_{k=1}^K \lambda_k \tilde{v}_k = \frac{z_0 - z_K}{\Lambda_K} \qquad \tilde{z}_K^a := \frac{1}{\Lambda_K}\sum_{k=1}^K \lambda_k \tilde{z}_k$$

$$\varepsilon_K^a := \frac{1}{\Lambda_K}\sum_{k=1}^K \lambda_k \left[\langle \tilde{v}_k, \tilde{z}_k - \tilde{z}_K^a\rangle + \varepsilon_k\right] \geq 0$$

and

$$\Lambda_k := \sum_{k=1}^K \lambda_k$$

Proof: Let $(z,v) \in \operatorname{gr} T$ be given. The HPE inclusion $\tilde{v}_k \in T^{\varepsilon_k}(\tilde{z}_k)$ implies that

$$\langle \tilde{v}_k - v, \tilde{z}_k - z\rangle \geq -\varepsilon_k$$

or equivalently,

$$\langle \tilde{v}_k, \tilde{z}_k - z\rangle + \varepsilon_k \geq \langle v, \tilde{z}_k - z\rangle$$

Hence

$$\frac{1}{\Lambda_K}\sum_{k=1}^K \lambda_k \left[\langle \tilde{v}_k, \tilde{z}_k - z\rangle + \varepsilon_k\right] \geq \frac{1}{\Lambda_K}\sum_{k=1}^K \lambda_k \langle v, \tilde{z}_k - z\rangle = \langle v, \tilde{z}_K^a - z\rangle$$

where the equality is due to the definition of $\tilde{z}_K^a$. Now let

$$\Gamma_K(z) := \frac{1}{\Lambda_K}\sum_{k=1}^K \lambda_k \left[\langle \tilde{z}_k - z, \tilde{v}_k\rangle + \varepsilon_K\right]$$

Then

$$\Gamma_K(z) \geq \langle v, \tilde{z}_K^a - z\rangle$$

Note that $\Gamma_K(\cdot)$ is an affine function such that

$$\nabla \Gamma_K = -\tilde{v}_K^a, \quad \Gamma_K(\tilde{z}_K^a) = \varepsilon_K^a$$

where these two identities follows from the definitions of $\tilde{v}_K^a$ and ε_K^a. Thus,

$$\begin{aligned}\Gamma_K(z) &= \Gamma_K(\tilde{z}_K^a) + \langle -\nabla\Gamma_K, \tilde{z}_K^a - z\rangle \\ &= \varepsilon_K^a + \langle \tilde{v}_K^a, \tilde{z}_K^a - z\rangle\end{aligned}$$

We then conclude that

$$\varepsilon_K^a + \langle \tilde{v}_K^a, \tilde{z}_K^a - z\rangle \geq \langle v, \tilde{z}_K^a - z\rangle$$

Since this holds for every $(z,v) \in \operatorname{gr} T$, we conclude that (1) holds. The fact that $\varepsilon_K^a \geq 0$ follows from (1) and the Technical Lemma above.

Lemma 7: For every $K \geq 1$, we have

$$\|\tilde{v}_K^a\| \leq \frac{2d_0}{\Lambda_K}$$

Proof: Have

$$\tilde{v}_K^a = \frac{1}{\Lambda_K} \sum_{k=1}^{K} \lambda_k \tilde{v}_k = \frac{1}{\Lambda_K} \sum_{k=1}^{K} (z_{k-1} - z_k) = \frac{z_0 - z_K}{\Lambda_K}$$

Hence,

$$\begin{aligned}
\|\tilde{v}_K^a\| = \frac{\|z_0 - z_K\|}{\Lambda_K} &\leq \frac{\|z_0 - z_* + z_* - z_K\|}{\Lambda_K} \\
&\leq \frac{\|z_0 - z_*\| + \|z_K - z_*\|}{\Lambda_K} \\
&\leq \frac{2\|z_0 - z_*\|}{\Lambda_K} = \frac{2d_0}{\Lambda_K}
\end{aligned}$$

Lemma 8: For every $k \geq 1$ and $z \in \mathbb{R}^n$, have

$$\|z - z_{k-1}\|^2 - \|z - z_k\|^2 \geq 2\lambda_k \gamma_k(z) + (1 - \sigma^2) \|\tilde{z}_k - z_{k-1}\|^2$$

where

$$\gamma_k(z) := \langle \tilde{z}_k - z, \tilde{v}_k \rangle + \varepsilon_k$$

Proof: By Lemma 1 and the definition of $\gamma_k(\cdot)$, have

$$\begin{aligned}
&\|z - z_{k-1}\|^2 - \|z - z_k\|^2 \\
&= 2\lambda_k \langle \tilde{z}_k - z, \tilde{v}_k \rangle + \|\tilde{z}_k - z_{k-1}\|^2 - \|\tilde{z}_k - z_k\|^2 \\
&= 2\lambda_k \gamma_k(z) + \|\tilde{z}_k - z_{k-1}\|^2 - \|\tilde{z}_k - z_k\|^2 - 2\lambda_k \varepsilon_k \\
&\geq 2\lambda_k \gamma_k(z) + (1 - \sigma^2) \|\tilde{z}_k - z_{k-1}\|^2
\end{aligned}$$

where the inequality is due to the HPE inequality condition.

Lemma 9: For every $K \geq 1$ and $z \in \mathbb{R}^n$, have

$$\|z - z_0\|^2 - \|z - z_K\|^2 \geq 2\Lambda_K \Gamma_K(z) + (1 - \sigma^2) \sum_{k=1}^{K} \|\tilde{z}_k - z_{k-1}\|^2$$

where

$$\Gamma_K(z) := \frac{1}{\Lambda_K} \sum_{k=1}^{K} \lambda_k \gamma_k(z) = \frac{1}{\Lambda_K} \sum_{k=1}^{K} \lambda_k \left[\langle \tilde{z}_k - z, \tilde{v}_k \rangle + \varepsilon_k\right]$$

Proof: This result follows from the definition of $\Gamma_k(\cdot)$ and by adding the inequality of Lemma 8 from $k = 1$ to $k = k$

Lemma 10: For every $K \geq 1$ and $z \in \mathbb{R}^n$, have

$$\|\tilde{z}_K^a - z_0\|^2 - \|\tilde{z}_K^a - z_K\|^2 \geq 2\Lambda_K \varepsilon_K^a + (1 - \sigma^2) \sum_{k=1}^{K} \|\tilde{z}_k - z_{k-1}\|^2$$

Proof: This result follows from the fact that

$$\varepsilon_K^a = \Gamma_K(\tilde{z}_K^a)$$

and the previous lemma with $z = \tilde{z}_K^a$

Lemma 11: For every $K \geq 1$ and $z \in \mathbb{R}^n$, have

$$\varepsilon_K^a \leq \left(4 + \frac{\sigma^2}{1-\sigma^2}\right) \frac{d_0^2}{\Lambda_K}$$

Proof: The previous lemma implies that

$$\varepsilon_K^a \leq \frac{\|\tilde{z}_K^a - z_0\|^2}{2\Lambda_K}$$

Now,

$$\begin{aligned}
\|\tilde{z}_K^a - z_0\|^2 &\leq \max\left\{\|\tilde{z}_K - z_0\|^2 : k = 1, \ldots, K\right\} \\
&\leq \max 2\left\{\|\tilde{z}_k - z_k\|^2 + \|z_k - z_0\|^2 : k = 1, \ldots, K\right\} \\
&\leq 2\max\left\{\|\tilde{z}_k - z_k\|^2 + 4d_0^2 : k = 1, \ldots, K\right\} \\
&\leq 8d_0^2 + 2\sigma^2 \max\left\{\|\tilde{z}_k - z_{k-1}\|^2 : k = 1, \ldots, K\right\} \\
&\leq 8d_0^2 + 2\sigma^2 \frac{d_0^2}{1-\sigma^2} = 2\left(4 + \frac{\sigma^2}{1-\sigma^2}\right) d_0^2
\end{aligned}$$

2 Generalized HPE framework (overview)

Consider the MIP

$$0 \in T(z)$$

where $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is maximal monotone

Assume that for some semi-norm in $\mathbb{R}^n$, the following conditions hold:

1) $T^{-1}(0) \neq \emptyset$
2) there exists $m, M > 0$ such that for every $z, z' \in \mathbb{R}^n$, have

$$(dw)_z(z') \geq \frac{m}{2}\|z' - z\|^2$$
$$\|\nabla w(z') - \nabla w(z)\|^* \leq M\,\|z' - z\|$$

where

$$\|\cdot\|^* := \sup\{\langle \cdot, v\rangle : \|v\| \leq 1\}$$

Condition 2) implies that

$$\frac{m}{2}\|z - z'\|^2 \leq (dw)_z(z') \leq \frac{M}{2}\|z - z'\|^2 \quad \forall z, z' \in Z \tag{2}$$

Example: If

$$w(\cdot) = (1/2)\|\cdot\|_Q^2$$

where Q is a self-adjoint positive semidefinite linear operator, then w satisfies condition 2 with $(m, M) = (1, 1)$.

Proposition 2.1 *If $Q : \mathbb{R}^n \to \mathbb{R}^n$ is a self-adjoint positive semidefinite linear operator, then the semi-norm*

$$\|\cdot\| := \langle \cdot, Q(\cdot)\rangle^{1/2}$$

satisfies the following statements:

(a) $\operatorname{dom}\|\cdot\|^* = \operatorname{Im}(Q)$ *and*

$$\|Qz\|^* = \|z\| \quad \forall z \in \mathbb{R}^n$$

(b) if Q is invertible, then

$$\|z\|^* = \langle Q^{-1}z, z\rangle^{1/2} \quad \forall z \in \mathbb{R}^n$$

Algorithm 1 (Inexact PPM framework with Bregman distance)

0. Let $z_0 \in \mathbb{R}^n$ and $\sigma \in [0, 1]$ be given.
1. Find $\lambda > 0$ and $(z, \tilde{z}, \varepsilon) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}_+$ such that

$$\frac{\nabla w(z_0) - \nabla w(z)}{\lambda} \in T^{\varepsilon}(\tilde{z}) \tag{3}$$

$$(dw)_z(\tilde{z}) + \lambda\varepsilon \leq \sigma(dw)_{z_0}(\tilde{z}) \tag{4}$$

2. Set $z_0 \leftarrow z$ and go to step 1.

Hence, sequence-wise, we have for every $k \geq 1$ that

$$r_k := \frac{\nabla w(z_{k-1}) - \nabla w(z_k)}{\lambda_k} \in T^{\varepsilon_k}(\tilde{z}_k) \tag{5}$$

$$(dw)_z(\tilde{z}_k) + \lambda_k\varepsilon_k \leq \sigma(dw)_{z_{k-1}}(\tilde{z}_k) \tag{6}$$

Proposition 2.2 *(***Pointwise***)* *Assume that $\sigma < 1$ and $\lambda_k \geq \underline{\lambda}$ for every $k \geq 1$. Then, for every $k \geq 1$, there exists $i \leq k$ such that*

$$\|r_i\|^* \leq \frac{2M}{\underline{\lambda}\sqrt{mk}}\sqrt{\frac{(1+\sigma)(dw)_0}{1-\sigma}} = \mathcal{O}\left(\frac{1}{\underline{\lambda}\sqrt{k}}\right)$$

and

$$\varepsilon_i \leq \frac{(1+\sigma)(dw)_0}{(1-\sigma)\underline{\lambda}k} = \mathcal{O}\left(\frac{1}{\underline{\lambda}k}\right)$$

where r_i is as in (5) *and*

$$(dw)_0 := \inf\left\{(dw)_{z_0}(z_*) : z_* \in T^{-1}(0)\right\}$$

For $k \geq 1$, define $\Lambda_k := \sum_{i=1}^k \lambda_i$ and the ergodic iterate $(\tilde{z}_k^a, r_k^a, \varepsilon_k^a)$ as

$$\tilde{z}_k^a = \frac{1}{\Lambda_k}\sum_{i=1}^k \lambda_i \tilde{z}_i, \quad r_k^a := \frac{1}{\Lambda_k}\sum_{i=1}^k \lambda_i r_i, \quad \varepsilon_k^a := \frac{1}{\Lambda_k}\sum_{i=1}^k \lambda_i \left(\varepsilon_i + \langle r_i, \tilde{z}_i - \tilde{z}_k^a\rangle\right). \quad (7)$$

Theorem 2.3 (Ergodic convergence of the NE-HPE) *For every $k \geq 1$, we have*

$$r_k^a \in T^{\varepsilon_k^a}(\tilde{z}_k^a)$$

and

$$\|r_k^a\|^* \leq \frac{2\sqrt{2}M(dw_0)^{1/2}}{\sqrt{m}\Lambda_k}, \quad \varepsilon_k^a \leq \left(\frac{3M}{m}\right)\left[\frac{3(dw)_0 + \sigma\theta_k}{\Lambda_k}\right].$$

where

$$\theta_k := \max_{i=1,\ldots,k} (dw)_{z_{i-1}}(\tilde{z}_i). \quad (8)$$

Moreover, the sequence $\{\theta_k\}$ is bounded under either one of the following situations:

(a) $\sigma < 1$, *in which case*

$$\theta_k \leq \frac{(dw)_0}{1-\sigma}; \quad (9)$$

(b) $\operatorname{Dom} T$ *is bounded, in which case*

$$\theta_k \leq \frac{2M}{m}[(dw)_0 + D]$$

where

$$D := \sup\{\min\{(dw)_y(y'), (dw)_{y'}(y)\} : y, y' \in \operatorname{Dom} T\}$$