8. Preliminaries

For any $x, y \in \mathbb{R}^d$, write $\langle x, y\rangle = x^T y$ for the inner product. We say a function $f : \mathbb{R}^d \to \mathbb{R} \cup \{\infty\}$ is convex if

$$f(\theta x + (1-\theta)y) \le \theta f(x) + (1-\theta) f(y)$$

for any $x, y \in \mathbb{R}^d$ and $\theta \in [0, 1]$. A convex function is closed if it is lower semi-continuous and proper if it is finite somwhere. We say f is μ-strongly convex for $\mu > 0$ if $f(x) - (\mu/2)\|x\|^2$ is a convex function. Given a convex function $f : \mathbb{R}^d \to \mathbb{R} \cup \{\infty\}$ and $\alpha > 0$, define its proximal operator $\mathrm{Prox}_f : \mathbb{R}^d \to \mathbb{R}^d$ as

$$\mathrm{Prox}_{\alpha f}(z) = \operatorname*{argmin}_{x\in\mathbb{R}^d} \left\{\alpha f(x) + (1/2)\|x - z\|^2\right\}.$$

When f is convex, closed, and proper, the argmin uniquely exists, and therefore Prox_f is well-defined. An mapping $T : \mathbb{R}^d \to \mathbb{R}^d$ is L-Lipschitz if

$$\|T(x) - T(y)\| \le L\|x - y\|$$

for all $x, y, \in \mathbb{R}^d$. If T is L-Lipschitz with $L \le 1$, we say T is nonexpansive. If T is L-Lipschitz with $L < 1$, we say T is a contraction. A mapping $T : \mathbb{R}^d \to \mathbb{R}^d$ is θ-averaged for $\theta \in (0, 1)$, if it is nonexpansive and if

$$T = \theta R + (1-\theta)I,$$

where $R : \mathbb{R}^d \to \mathbb{R}^d$ is another nonexpansive mapping.

Lemma 4 (Proposition 4.35 of (Bauschke & Combettes, 2017))**.** *$T : \mathbb{R}^d \to \mathbb{R}^d$ is θ-averaged if and only if*

$$\|T(x) - T(y)\|^2 + (1-2\theta)\|x - y\|^2 \le 2(1-\theta)\langle T(x) - T(y), x - y\rangle$$

for all $x, y \in \mathbb{R}^d$.

Lemma 5 ((Ogura & Yamada, 2002; Combettes & Yamada, 2015))**.** *Assume $T_1 : \mathbb{R}^d \to \mathbb{R}^d$ and $T_2 : \mathbb{R}^d \to \mathbb{R}^d$ are θ_1 and θ_2-averaged, respectively. Then T_1T_2 is $\frac{\theta_1+\theta_2-2\theta_1\theta_2}{1-\theta_1\theta_2}$-averaged.*

Lemma 6. *Let $T : \mathbb{R}^d \to \mathbb{R}^d$. $-T$ is θ-averaged if and only if $T \circ (-I)$ is θ-averaged.*

Proof. The lemma follows from the fact that

$$T \circ (-I) = \theta R + (1-\theta)I \quad \Leftrightarrow \quad -T = \theta(-R) \circ (-I) + (1-\theta)I$$

for some nonexpansive R and that nonexpansiveness of R and implies nonexpansivenes of $-R \circ (-I)$. □

Lemma 7 ((Taylor et al., 2018))**.** *Assume f is μ-strongly convex and ∇f is L-Lipschitz. Then for any $x, y \in \mathbb{R}^d$, we have*

$$\|(I - \alpha\nabla f)(x) - (I - \alpha\nabla f)(y)\| \le \max\{|1 - \alpha\mu|, |1 - \alpha L|\}\|x - y\|.$$

Lemma 8 (Proposition 5.4 of (Giselsson, 2017))**.** *Assume f is μ-strongly convex, closed, and proper. Then*

$$-(2\mathrm{Prox}_{\alpha f} - I)$$

is $\frac{1}{1+\alpha\mu}$-averaged.

References. The notion of proximal operator and its well-definedness were first presented in (Moreau, 1965). The notion of averaged mappings were first introduced in (Bailion et al., 1978). The idea of Lemma 6 relates to "negatively averaged" operators from (Giselsson, 2017). Lemma 7 is proved in a weaker form as Theorem 3 of (Polyak, 1987) and in Section 5.1 of (Ryu & Boyd, 2016). Lemma 7 as stated is proved as Theorem 2.1 in (Taylor et al., 2018).

9. Proofs of main results

9.1. Equivalence of PNP-DRS and PNP-ADMM

We show the standard steps that establish equivalence of PNP-DRS and PNP-ADMM. Starting from PNP-DRS, we substitute $z^k = x^k + u^k$ to get

$$\begin{aligned} x^{k+1/2} &= \mathrm{Prox}_{\alpha f}(x^k + u^k) \\ x^{k+1} &= H_\sigma(x^{k+1/2} - (u^k + x^k - x^{k+1/2})) \\ u^{k+1} &= u^k + x^k - x^{k+1/2}. \end{aligned}$$

We reorder the iterations to get the correct dependency

$$\begin{aligned} x^{k+1/2} &= \mathrm{Prox}_{\alpha f}(x^k + u^k) \\ u^{k+1} &= u^k + x^k - x^{k+1/2} \\ x^{k+1} &= H_\sigma(x^{k+1/2} - u^{k+1}). \end{aligned}$$

We label $\tilde{y}^{k+1} = x^{k+1/2}$ and $\tilde{x}^{k+1} = x^k$

$$\begin{aligned} \tilde{x}^{k+1} &= H_\sigma(\tilde{y}^k - u^k) \\ \tilde{y}^{k+1} &= \mathrm{Prox}_{\alpha f}(\tilde{x}^{k+1} + u^k) \\ u^{k+1} &= u^k + \tilde{x}^{k+1} - \tilde{y}^{k+1}, \end{aligned}$$

and we get PNP-ADMM.

9.2. Convergence analysis

Lemma 9. *$H_\sigma : \mathbb{R}^d \to \mathbb{R}^d$ satisfies Assumption* (A) *if and only if*

$$\frac{1}{1+\varepsilon} H_\sigma$$

is nonexpansive and $\frac{\varepsilon}{1+\varepsilon}$-averaged.

Proof. Define $\theta = \frac{\varepsilon}{1+\varepsilon}$, which means $\varepsilon = \frac{\theta}{1-\theta}$. Clearly, $\theta \in [0, 1)$. Define $G = \frac{1}{1+\varepsilon} H_\sigma$, which means $H_\sigma = \frac{1}{1+\theta} G$. Then

$$\begin{aligned} &\underbrace{\|(H_\sigma - I)(x) - (H_\sigma - I)(y)\|^2 - \frac{\theta^2}{(1-\theta)^2}\|x-y\|^2}_{\text{(TERM A)}} \\ &\quad = \frac{1}{(1-\theta)^2}\|G(x) - G(y)\|^2 + \left(1 - \frac{\theta^2}{(1-\theta)^2}\right)\|x-y\|^2 - \frac{2}{1-\theta}\langle G(x) - G(y), x - y\rangle \\ &\quad = \frac{1}{(1-\theta)^2}\Bigg(\underbrace{\|G(x) - G(y)\|^2 + (1-2\theta)\|x-y\|^2 - 2(1-\theta)\langle G(x) - G(y), x-y\rangle}_{\text{(TERM B)}}\Bigg). \end{aligned}$$

Remember that Assumption (A) corresponds to (TERM A) ≤ 0 for all $x, y \in \mathbb{R}^d$. This is equivalent to (TERM B) ≤ 0 for all $x, y \in \mathbb{R}^d$, which corresponds to G being θ-averaged by Lemma 4. □

Lemma 10. *$H_\sigma : \mathbb{R}^d \to \mathbb{R}^d$ satisfies Assumption* (A) *if and only if*

$$\frac{1}{1+2\varepsilon}(2H_\sigma - I)$$

is nonexpansive and $\frac{2\varepsilon}{1+2\varepsilon}$-averaged.

Proof. Define $\theta = \frac{2\varepsilon}{1+2\varepsilon}$, which means $\varepsilon = \frac{\theta}{2(1-\theta)}$. Clearly, $\theta \in [0, 1)$. Define $G = \frac{1}{1+2\varepsilon}(2H_\sigma - I)$, which means $H_\sigma = \frac{1}{2(1-\theta)}G + \frac{1}{2}I$. Then

$$\begin{aligned}
&\underbrace{\|(H_\sigma - I)(x) - (H_\sigma - I)(y)\|^2 - \frac{\theta^2}{4(1-\theta)^2}\|x-y\|^2}_{\text{(TERM A)}} \\
&\quad = \frac{1}{4(1-\theta)^2}\|G(x)-G(y)\|^2 + \left(\frac{1}{4} - \frac{\theta^2}{4(1-\theta)^2}\right)\|x-y\|^2 - \frac{1}{2(1-\theta)}\langle G(x)-G(y), x-y\rangle \\
&\quad = \frac{1}{4(1-\theta)^2}\Bigg(\underbrace{\|G(x)-G(y)\|^2 + (1-2\theta)\|x-y\|^2 - 2(1-\theta)\langle G(x)-G(y), x-y\rangle}_{\text{(TERM B)}}\Bigg).
\end{aligned}$$

Remember that Assumption (A) corresponds to (TERM A) ≤ 0 for all $x, y \in \mathbb{R}^d$. This is equivalent to (TERM B) ≤ 0 for all $x, y \in \mathbb{R}^d$, which corresponds to G being θ-averaged by Lemma 4. □

Proof of Theorem 1. In general, if operators T_1 and T_2 are L_1 and L_2-Lipschitz, then the composition T_1T_2 is (L_1L_2)-Lipschitz. By Lemma 7, $I - \alpha\nabla f$ is $\max\{|1-\alpha\mu|, |1-\alpha L|\}$-Lipschitz. By Lemma 9, H_σ is $(1+\varepsilon)$-Lipschitz. The first part of the theorem following from composing the Lipschitz constants. The restrictions on α and ε follow from basic algebra. □

Proof of Theorem 2. By Lemma 8,

$$-(2\mathrm{Prox}_{\alpha f} - I)$$

is $\frac{1}{1+\alpha\mu}$-averaged, and this implies

$$(2\mathrm{Prox}_{\alpha f} - I) \circ (-I)$$

is also $\frac{1}{1+\alpha\mu}$-averaged, by Lemma 6. By Lemma 10,

$$\frac{1}{1+2\varepsilon}(2H_\sigma - I)$$

is $\frac{2\varepsilon}{1+2\varepsilon}$-averaged. Therefore,

$$\frac{1}{1+2\varepsilon}(2H_\sigma - I)(2\mathrm{Prox}_{\alpha f} - I) \circ (-I)$$

is $\frac{1+2\varepsilon\alpha\mu}{1+\alpha\mu+2\varepsilon\alpha\mu}$-averaged by Lemma 5, and this implies

$$-\frac{1}{1+2\varepsilon}(2H_\sigma - I)(2\mathrm{Prox}_{\alpha f} - I)$$

is also $\frac{1+2\varepsilon\alpha\mu}{1+\alpha\mu+2\varepsilon\alpha\mu}$-averaged, by Lemma 6.

Using the definition of averagedness, we can write

$$(2H_\sigma - I)(2\mathrm{Prox}_{\alpha f} - I) = -(1+2\varepsilon)\left(\frac{\alpha\mu}{1+\alpha\mu+2\varepsilon\alpha\mu}I + \frac{1+2\varepsilon\alpha\mu}{1+\alpha\mu+2\varepsilon\alpha\mu}R\right)$$

where R is a nonexpansive operator. Plugging this into the PNP-DRS operator, we get

$$\begin{aligned}
T &= \frac{1}{2}I - \frac{1}{2}(1+2\varepsilon)\left(\frac{\alpha\mu}{1+\alpha\mu+2\varepsilon\alpha\mu}I + \frac{1+2\varepsilon\alpha\mu}{1+\alpha\mu+2\varepsilon\alpha\mu}R\right) \\
&= \underbrace{\frac{1}{2(1+\alpha\mu+2\varepsilon\alpha\mu)}}_{=A} I - \underbrace{\frac{(1+2\varepsilon\alpha\mu)(1+2\varepsilon)}{2(1+\alpha\mu+2\varepsilon\alpha\mu)}}_{=B} R,
\end{aligned} \tag{1}$$

where define the coefficients A and B for simplicity. Clearly, $A > 0$ and $B > 0$. Then we have

$$\begin{aligned}\|Tx - Ty\|^2 &= A^2\|x-y\|^2 + B^2\|R(x)-R(y)\|^2 - 2\langle A(x-y), B(R(x)-R(y))\rangle \\ &\le A^2\left(1+\frac{1}{\delta}\right)\|x-y\|^2 + B^2\left(1+\delta\right)\|R(x)-R(y)\|^2 \\ &\le \left(A^2\left(1+\frac{1}{\delta}\right) + B^2\left(1+\delta\right)\right)\|x-y\|^2\end{aligned}$$

for any $\delta > 0$. The first line follows from plugging in (1). The second line follows from applying Young's inequality to the inner product. The third line follows from nonexpansiveness of R.

Finally, we optimize the bound. It is a matter of simple calculus to see

$$\min_{\delta>0}\left\{A^2\left(1+\frac{1}{\delta}\right) + B^2\left(1+\delta\right)\right\} = (A+B)^2.$$

Plugging this in, we get

$$\|Tx-Ty\|^2 \le (A+B)^2\|x-y\|^2 = \left(\frac{1+\varepsilon+\varepsilon\alpha\mu+2\varepsilon^2\alpha\mu}{1+\alpha\mu+2\varepsilon\alpha\mu}\right)^2\|x-y\|^2,$$

which is the first part of the theorem.

The restrictions on α and ε follow from basic algebra. □

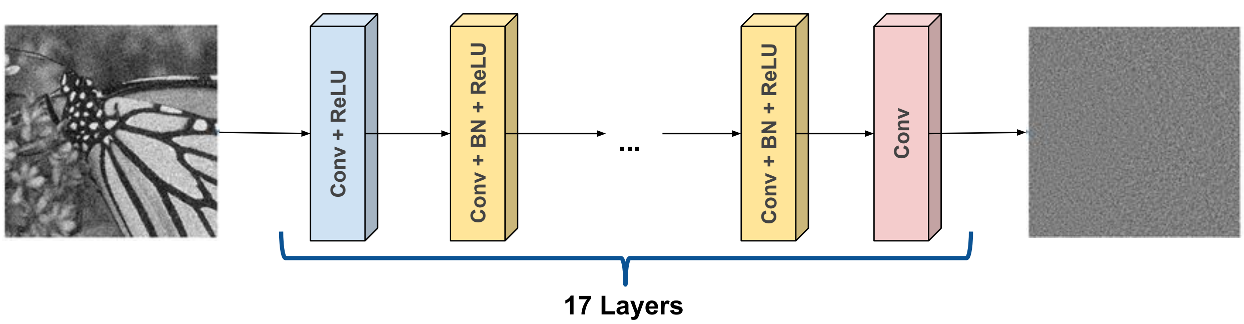

Figure 3. DnCNN Network Architecture

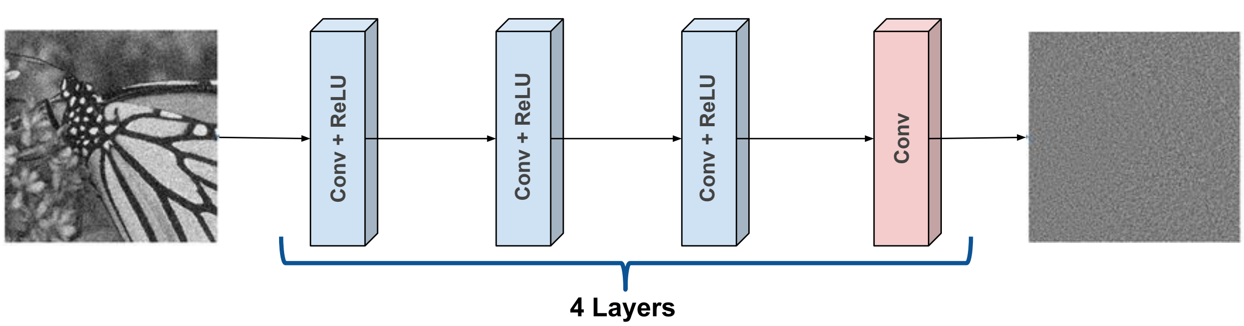

Figure 4. SimpleCNN Network Architecture