Notes on Learned Proximal Networks

- Let $\underline{x} \in \mathbb{R}^n$, $A: \mathbb{R}^m \to \mathbb{R}^n$, and consider the inverse problem
$$\underline{y} = \underline{A}(\underline{x}) + \underline{V}, \tag{1}$$
where $\underline{V}$ is some noise/nuisance term. Our goal is to recover a solution $\hat{\underline{x}} \in \mathbb{R}^n$ to this problem that approximates the real signal $\underline{x}$ or even recover it.

- This is an ill-posed inverse problem. The inverse problem may have an infinite number of solutions all approximating the signal ... To overcome this, we need to regularize the problem.

- One approach: Use a prior to regularise the problem. This is a function $R: \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ that:

    i) Promote certain desirable properties in $\underline{x}$ or
    ii) Encapsulate the knowledge we have about $\underline{x}$ or
    iii) Promote a solution likely under the distribution of $\underline{x}$,

    or any combinations of i) – iii).

⚠ Using a prior function R does NOT mean necessarily that $\underline{x}$ is sampled from a distribution $p \sim e^{-R(\cdot)}$ or is likely a sample from $\sim e^{-R(\cdot)}$ or that samples of $\sim e^{-R(\cdot)}$ are similar to $\underline{x}$. There are counterexamples to this. (E.g., $R = \|\cdot\|_1$.)

- Assume now that the noise $\underline{v}$ is Gaussian. Then an appropriate data fidelity term is the quadratic norm $\frac{1}{2}\|\cdot\|_2^2$.

- We can recover a solution with the data fidelity term and the prior R by minimizing their weighted sum:

$$\hat{\underline{x}}_R(\underline{y}, t) \in \underset{\underline{x} \in \operatorname{dom} R}{\operatorname{argmin}} \ \frac{1}{2t}\|\underline{y} - \underline{A}(\underline{x})\|_2^2 + R(\underline{x}), \qquad (2)$$

where $t > 0$ is a hyperparameter.

(Note: We minimize over dom R to allow for the prior R to take the extended value $+\infty$ on some subset of $\mathbb{R}^n$. This allows for, e.g., priors defined on bounded domains.)

~~~

- Key to our work is the proximal operator of prior R, which we will denote by prox_R:

$$\text{prox}_R(\underline{y}) \in \underset{\underline{x} \in \operatorname{dom} R}{\operatorname{argmin}} \ \frac{1}{2}\|\underline{y} - \underline{x}\|_2^2 + R(\underline{x}). \qquad (3)$$

When R is proper (i.e., not identically $+\infty$ and nowhere equal to $-\infty$), lower semicontinuous and convex, then prox_R is single-valued.

When R is non-convex, then we do not have uniqueness of solutions in (3)...
~~~

- Following Gribonval and Nikolova (2020), we define the proximal operator of R as a selection over the solutions of (3).

- More precisely, a function $f : \mathrm{dom}\, R \to \mathbb{R}^n$ is a proximity operator of a prior $R : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ if

$$f(y) \in \underset{x \in \mathrm{dom}\, R}{\mathrm{argmin}} \ \frac{1}{2}\|y - x\|_2^2 + R(x) \tag{4}$$

for each $y \in \mathrm{dom}\, R$.

- For short, we will write (4) as $f(y) \in \mathrm{prox}_R(y)$.

- Theorem 1 + Corollary 1 from Gribonval and Nikolova (2020) yield the following:

Let $\mathrm{dom}\, R$ be non-empty and open. Then f is a (continuous) proximal operator of R if and only if there exists a convex, differentiable function Ψ on $\mathrm{dom}\, R$ such that

$$f(y) = \nabla_y \Psi(y) \quad \text{for each } y \in \mathrm{dom}\, R.$$

Moreover, we have the identity

$$\Psi(y) + \left(R(\nabla_y \Psi(y)) + \frac{1}{2}\|\nabla_y \Psi(y)\|_2^2\right) = \langle y, \nabla_y \Psi(y)\rangle. \tag{5}$$