Mathematical methods for Image Processing

François Malgouyres

Institut de Mathématiques de Toulouse, France

invitation by
Jidesh P., NITK Surathkal

funding
Global Initiative on Academic Network

Oct. 23–27

Plan

The non-smooth problem

We consider the minimization problem

$$w^* \in \operatorname{Argmin}_{w \in W} \mathcal{E}(w)$$

with

$$\mathcal{E}(w) = E(w) + R(w) \qquad , \text{ for all } w \in W,$$

where

- W is a Euclidean space
- E is convex, coercive, differentiable with a Lipschitz gradient
- R is lower-semi-continuous, proper, convex and **simple**.

Definition (proximal operator and simple)

We say R is **simple** *if there is a simple way to compute*

$$\operatorname{prox}_R^t(w') = \operatorname{Argmin}_{w \in W} \frac{t}{2}\|w - w'\|_2^2 + R(w).$$

(e.g. It is given in a closed form expression or computed by a fast algorithm.)

Example 1 : R is a characteristic function

Let $C \subset W$ be a non-empty closed set :

$$R(w) = \chi_C(w) = \begin{cases} 0 & \text{, if } w \in C \\ +\infty & \text{, otherwise.} \end{cases}$$

Then,

$$\begin{aligned} \mathrm{prox}_R^t(w') &= \mathrm{Argmin}_{w\in W} \frac{t}{2}\|w - w'\|_2^2 + R(w) \\ &= \mathrm{Argmin}_{w\in C} \|w - w'\|_2^2 \end{aligned}$$

$\mathrm{prox}_R^t(w')$ is the projection onto C.

Usually easy to compute when (for instance) C is

- an ℓ^1, ℓ^2 or ℓ^∞ ball
- an affine space.

Example 2 : R is $\|.\|_1$

If

$$R(w) = \|w\|_1 = \sum_i |w_i|$$

we have

$$\begin{aligned} \mathrm{prox}^L_{\|.\|_1}(w') &= \mathrm{Argmin}_{w \in W} \frac{L}{2}\|w - w'\|_2^2 + \|w\|_1, \\ &= \mathrm{Argmin}_{w \in W} \sum_i \left(\frac{L}{2}(w_i - w'_i)^2 + |w_i| \right). \end{aligned} \tag{1}$$

The i^{th} entry of $\mathrm{prox}^L_{\|.\|_1}(w')$ is

$$\mathrm{prox}^L_{\|.\|_1}(w')_i = \mathrm{Argmin}_{t \in \mathbb{R}} \frac{L}{2}(t - w'_i)^2 + |t|.$$

Example 2 : R is $\|.\|_1$

The i^{th} entry of $\text{prox}^L_{\|.\|_1}(w')$ is

$$\text{prox}^L_{\|.\|_1}(w')_i = \text{Argmin}_{t\in\mathbb{R}} \frac{L}{2}(t - w'_i)^2 + |t|.$$

Proof : Let $v_i = \text{Argmin}_{t\in\mathbb{R}} \frac{L}{2}(t - w'_i)^2 + |t|$ and $v = (v_i)_i$. Since for every i and every $w \in W$

$$\frac{L}{2}(v_i - w'_i)^2 + |v_i| \leq \frac{L}{2}(w_i - w'_i)^2 + |w_i|,$$

we have

$$\begin{array}{rcl} \sum_i \left(\frac{L}{2}(v_i - w'_i)^2 + |v_i|\right) & \leq & \sum_i \left(\frac{L}{2}(w_i - w'_i)^2 + |w_i|\right) \\ \frac{L}{2}\|v - w'\|_2^2 + \|v\|_1 & \leq & \frac{L}{2}\|w - w'\|_2^2 + \|w\|_1 \end{array}$$

Therefore $\text{prox}^L_{\|.\|_1}(w') = v$.

Example 2 : R is $\|.\|_1$

$\mathrm{prox}^L_{\|.\|_1}(w')_i$ is obtained by a soft thresholding

$$\mathrm{prox}^L_{\|.\|_1}(w')_i = \begin{cases} w'_i - \frac{1}{L} & \text{, si } w'_i > \frac{1}{L}. \\ 0 & \text{, si } -\frac{1}{L} \leq w'_i \leq \frac{1}{L}, \\ w'_i + \frac{1}{L} & \text{, si } w'_i < -\frac{1}{L}, \end{cases}$$

Proof : We remind that $\mathrm{prox}^L_{\|.\|_1}(w')_i = \mathrm{Argmin}_{t\in\mathbb{R}} \frac{L}{2}(t - w'_i)^2 + |t|$ and distinguish three cases
Case 1 :

$$\begin{aligned} \mathrm{prox}^L_{\|.\|_1}(w')_i > 0 \quad & \Longleftrightarrow \quad \mathrm{prox}^L_{\|.\|_1}(w')_i > 0 \text{ and } L(\mathrm{prox}^L_{\|.\|_1}(w')_i - w'_i) + 1 = 0 \\ & \Longleftrightarrow \quad \mathrm{prox}^L_{\|.\|_1}(w')_i = w'_i - \frac{1}{L} \text{ and } w'_i > \frac{1}{L} \end{aligned}$$

Example 2 : R is $\|.\|_1$

$\mathrm{prox}^L_{\|.\|_1}(w')_i$ is obtained by a soft thresholding

$$\mathrm{prox}^L_{\|.\|_1}(w')_i = \begin{cases} w'_i - \frac{1}{L} & \text{, si } w'_i > \frac{1}{L}. \\ 0 & \text{, si } -\frac{1}{L} \leq w'_i \leq \frac{1}{L}, \\ w'_i + \frac{1}{L} & \text{, si } w'_i < -\frac{1}{L}, \end{cases}$$

Proof : We remind that $\mathrm{prox}^L_{\|.\|_1}(w')_i = \mathrm{Argmin}_{t\in\mathbb{R}} \frac{L}{2}(t-w'_i)^2 + |t|$ and distinguish three cases
Case 2 :

$$\begin{aligned} \mathrm{prox}^L_{\|.\|_1}(w')_i = 0 \quad & \Longleftrightarrow \quad \mathrm{prox}^L_{\|.\|_1}(w')_i = 0 \text{ and } L(\mathrm{prox}^L_{\|.\|_1}(w')_i - w'_i) \in [-1,1] \\ & \Longleftrightarrow \quad \mathrm{prox}^L_{\|.\|_1}(w')_i = 0 \text{ and } -\frac{1}{L} \leq w'_i \leq \frac{1}{L} \end{aligned}$$

Example 2 : R is $\|.\|_1$

$\text{prox}^L_{\|.\|_1}(w')_i$ is obtained by a soft thresholding

$$\text{prox}^L_{\|.\|_1}(w')_i = \begin{cases} w'_i - \frac{1}{L} & \text{, si } w'_i > \frac{1}{L}. \\ 0 & \text{, si } -\frac{1}{L} \leq w'_i \leq \frac{1}{L}, \\ w'_i + \frac{1}{L} & \text{, si } w'_i < -\frac{1}{L}, \end{cases}$$

Proof : We remind that $\text{prox}^L_{\|.\|_1}(w')_i = \text{Argmin}_{t\in\mathbb{R}} \frac{L}{2}(t - w'_i)^2 + |t|$ and distinguish three cases
Case 3 :

$$\begin{aligned} \text{prox}^L_{\|.\|_1}(w')_i < 0 \quad & \Longleftrightarrow \quad \text{prox}^L_{\|.\|_1}(w')_i < 0 \text{ and } L(\text{prox}^L_{\|.\|_1}(w')_i - w'_i) - 1 = 0 \\ & \Longleftrightarrow \quad \text{prox}^L_{\|.\|_1}(w')_i = w'_i + \frac{1}{L} \text{ and } w'_i < -\frac{1}{L} \end{aligned}$$

Example 3 : smooth case

If R is continuously differentiable

$$\mathrm{prox}_R^t(w') = \mathrm{Argmin}_{w \in W} \frac{t}{2}\|w - w'\|_2^2 + R(w)$$

satisfies

$$t\left(\mathrm{prox}_R^t(w') - w'\right) + \nabla R(\mathrm{prox}_R^t(w')) = 0.$$

therefore

$$\mathrm{prox}_R^t(w') = w' - \frac{1}{t}\nabla R(\mathrm{prox}_R^t(w')).$$

$\mathrm{prox}_R^t(w')$ **is an implicit gradient step with step-size** $\frac{1}{t}$**.**

The proximal gradient algorithm

Also known as "forward-backward algorithm", "implicit-explicit", "ISTA", "PALM" ...

Algorithm 2 Proximal gradient algorithm

Entry: Entry needed for computing E, ∇E and $\text{prox}_R^t(.)$
Output: Approximation of a minimizer of $\mathcal{E}$: w^*

```
Initialize $w$
While Not converged Do
    Compute $d = \nabla E(w)$
    Compute a step-size $t \geq 0$
    Update : $w \leftarrow \text{prox}_R^t(w - t\, d)$
End while
```

Convergence of the Proximal Gradient Algorithm

Theorem (Convergence of the Proximal Gradient algorithm)

We consider $\mathcal{E} = E + R$

- *Where* $E : W \longrightarrow \mathbb{R}$ *is convexe, coercive, differentiable, with a Lipschitz gradient*[a] *of constant* $L > 0$
- *Where* R *is lower semi-continuous, proper, convex and coercive.*

The sequence $(w_k)_{k\in\mathbb{N}}$ *generated by the Proximal gradient Algorithm for a step-size* $t < \frac{1}{L}$ *is such that*

- $(\mathcal{E}(w^k))_{k\in\mathbb{N}}$ *is non-increasing*
- *For any minimizer* w^* *of* $\mathcal{E}$

$$\mathcal{E}(w^k) - \mathcal{E}(w^*) \leq \frac{L}{2k}\|w^0 - w^*\|_2.$$

---

[a]

$$\forall w, w' \in W, \qquad \|\nabla E(w') - \nabla E(w)\| \leq L\|w' - w\|$$

Proof (Majorize-Minorize)

Lemma (A quadratique majorant)

We have for any w, w′ ∈ W

$$E(w') \leq E(w) + \langle \nabla E(w), w' - w \rangle + \frac{L}{2} \|w' - w\|_2^2.$$

Proof Using the second fundamental theorem of calculus, we have

$$E(w') = E(w) + \int_0^1 \langle \nabla E(tw + (1-t)w'), w' - w \rangle dt.$$

Therefore

$$\begin{aligned} & E(w') - E(w) - \langle \nabla E(w), w' - w \rangle \\ = \ & \int_0^1 \langle \nabla E(tw + (1-t)w') - \nabla E(w), w' - w \rangle dt, \end{aligned}$$

Proof (Majorize-Minorize)

Lemma (A quadratique majorant)

We have for any w, $w' \in W$

$$E(w') \leq E(w) + \langle \nabla E(w), w' - w \rangle + \frac{L}{2}\|w' - w\|_2^2.$$

End of the proof

$$\begin{aligned}
& E(w') - E(w) - \langle \nabla E(w), w' - w \rangle \\
= \quad & \int_0^1 \langle \nabla E(tw + (1-t)w') - \nabla E(w), w' - w \rangle dt, \\
\leq \quad & \int_0^1 \|\nabla E(tw + (1-t)w') - \nabla E(w)\|_2 \; \|w' - w\|_2 dt, \\
\leq \quad & \int_0^1 L\|tw + (1-t)w' - w\|_2 \; \|w' - w\|_2 dt, \\
= \quad & L\|w' - w\|_2^2 \int_0^1 (1-t) dt = \frac{L}{2}\|w' - w\|_2^2.
\end{aligned}$$

□

Proof (Majorize-Minorize)

We denote for $k \geq 1$ and $w \in W$,

$$F_k(w) = E(w^{k-1}) + \left\langle \nabla E(w^{k-1}), w - w^{k-1} \right\rangle + \frac{L}{2}\|w - w^{k-1}\|_2^2.$$

We have (using the previous Lemma)

$$E(w) \leq F_k(w). \tag{1}$$

Lemma (**Minorize)**

We have

$$w^k = \operatorname{Argmin}_{w \in W} F_k(w) + R(w). \tag{2}$$

We also have for all $w \in W$

$$F_k(w) + R(w) \geq F_k(w^k) + R(w^k) + \frac{L}{2}\|w - w^k\|_2^2. \tag{3}$$

Proof (Majorize-Minorize)

Proof of (2)**:** $w^k = \text{Argmin}_{w\in W} F_k(w) + R(w)$

$$
\begin{aligned}
w^k &= \text{prox}_R^L\left(w^{k-1} \; - \; \frac{1}{L}\,\nabla E(w^{k-1})\right), \\
&= \text{Argmin}_{w\in W} \; \frac{L}{2}\|w - w^{k-1} \; + \; \frac{1}{L}\,\nabla E(w^{k-1})\|_2^2 + R(w), \\
&= \text{Argmin}_{w\in W} \; \frac{1}{2L}\|\nabla E(w^{k-1})\|_2^2 \\
&\qquad\qquad + \langle w - w^{k-1}, \nabla E(w^{k-1})\rangle + \frac{L}{2}\|w - w^{k-1}\|_2^2 + R(w), \\
&= \text{Argmin}_{w\in W} \; F_k(w) + R(w).
\end{aligned}
$$

Proof (Majorize-Minorize)

Proof of (3) **:** $F_k(w) + R(w) \geq F_k(w^k) + R(w^k) + \frac{L}{2}\|w - w^k\|_2^2$

First notice that, for all $w \in W$

$$\begin{aligned}
& F_k(w) - F_k(w^k) \\
= \quad & E(w^{k-1}) + \left\langle \nabla E(w^{k-1}), w - w^{k-1} \right\rangle + \frac{L}{2}\|w - w^{k-1}\|_2^2 \\
& -\Big((E(w^{k-1}) + \left\langle \nabla E(w^{k-1}), w^k - w^{k-1} \right\rangle + \frac{L}{2}\|w^k - w^{k-1}\|_2^2 \Big), \\
= \quad & \left\langle \nabla E(w^{k-1}), w - w^k \right\rangle + \frac{L}{2}\left(\|w - w^k\|_2^2 + 2\left\langle w - w^k, w^k - w^{k-1} \right\rangle \right), \\
= \quad & \frac{L}{2}\|w - w^k\|_2^2 + \left\langle \nabla E(w^{k-1}) + L(w^k - w^{k-1}), w - w^k \right\rangle, \\
= \quad & \frac{L}{2}\|w - w^k\|_2^2 + \left\langle \nabla F_k(w^k), w - w^k \right\rangle.
\end{aligned}$$

Proof (Majorize-Minorize)

End of the proof of (3) **:** $F_k(w)+R(w)\geq F_k(w^k)+R(w^k)+\frac{L}{2}\|w-w^k\|_2^2$

$$F_k(w)+R(w)-F_k(w^k)-R(w^k)=\frac{L}{2}\|w-w^k\|_2^2+\left\langle\nabla F_k(w^k),w-w^k\right\rangle \\ +R(w)-R(w^k).$$

Moreover, since $w^k=\mathrm{Argmin}_{w\in W}\,F_k(w)+R(w)$,

$$0\in\partial(F_k+R)(w^k)=\nabla F_k(w^k)+\partial R(w^k),$$

we have

$$w^k\in\mathrm{Argmin}_{w\in W}\ \left\langle\nabla F_k(w^k),w-w^k\right\rangle\ +\ R(w).$$

Therefore

$$\left\langle\nabla F_k(w^k),w-w^k\right\rangle+R(w)\geq R(w^k)$$

and

$$F_k(w)+R(w)-F_k(w^k)-R(w^k)\geq\frac{L}{2}\|w-w^k\|_2^2.$$

□

Proof (Majorize-Minorize)

Let us resume to the proof of the main Theorem.

Let $w^* \in \operatorname{Argmin}_{w\in W} \mathcal{E}(w)$, we have using (1) ($E(w) \leq F_k(w)$), Lemma 2, and the convexity of E that

$$\begin{aligned}
\mathcal{E}(w^k) &\leq F_k(w^k) + R(w^k) \\
&\leq F_k(w^*) + R(w^*) - \frac{L}{2}\|w^* - w^k\|_2^2 \\
&= E(w^{k-1}) + \left\langle \nabla E(w^{k-1}), w^* - w^{k-1} \right\rangle + R(w^*) \\
&\qquad + \frac{L}{2}\|w^* - w^{k-1}\|_2^2 - \frac{L}{2}\|w^* - w^k\|_2^2 \\
&\leq \mathcal{E}(w^*) + \frac{L}{2}\left(\|w^* - w^{k-1}\|_2^2 - \|w^* - w^k\|_2^2\right)
\end{aligned}$$

Proof (Majorize-Minorize)

Using $E(w^k) \leq F_k(w^k)$ and $w^k = \text{Argmin}_{w \in W} F_k(w) + R(w)$, we have

$$\mathcal{E}(w^k) \leq F_k(w^k) + R(w^k) \leq F_k(w^{k-1}) + R(w^{k-1}) = \mathcal{E}(w^{k-1}).$$

In words $(\mathcal{E}(w^k))_{k \in \mathbb{N}}$ is non-increasing. We therefore have for all $k' \leq k$

$$\mathcal{E}(w^k) - \mathcal{E}(w^*) \leq \mathcal{E}(w^{k'}) - \mathcal{E}(w^*).$$

and therefore

$$\begin{aligned}
\mathcal{E}(w^k) - \mathcal{E}(w^*) &\leq \frac{1}{k} \sum_{k'=1}^{k} \Big(\mathcal{E}(w^{k'}) - \mathcal{E}(w^*) \Big) \\
&\leq \frac{L}{2k} \sum_{k'=1}^{k} \Big(\|w^* - w^{k'-1}\|_2^2 - \|w^* - w^{k'}\|_2^2 \Big) \\
&\leq \frac{L}{2k} \left(\|w^* - w^0\|_2^2 - \|w^* - w^k\|_2^2 \right) \\
&\leq \frac{L}{2k} \|w^* - w^0\|_2^2
\end{aligned}$$

☐

To go further

- Accelerated version exists (convergence in $O(\frac{1}{k^2})$) : FISTA (Beck-Teboulle)
- For other algorithm using the prox, see : Chambolle-Pock Algorithm, Douglas-Rachford algorithm, Proximal Point Algorithm
- Convergence proof including non-convex settings : PALM (Bolte-Sabach-Teboulle)
- Including a Stochastic setting : Chouzenoux-Pesquet-Reppeti