Plug-and-Play Image Restoration

Arthur Leclaire

Joint work with Samuel Hurault and Nicolas Papadakis

MVA Generative Modeling
March, 5th, 2024

Last Week

- We introduced Plug-and-Play (PnP) Image restoration.
- PnP are based on proximal splitting algorithms to minimize $f + \lambda g$ (with $f(x) = -\log p(y|x)$ and regularization $g(x)$)
- The main idea is to replace the regularization step by an off-the-shelf denoiser.
- For example, for Proximal Gradient Descent (PGD)

$$x_{k+1} = \mathrm{Prox}_{\tau\lambda g} \circ (\mathrm{Id} - \tau\nabla f)(x_k) \quad \text{becomes} \quad x_{k+1} = D_\sigma \circ (\mathrm{Id} - \tau\nabla f)(x_k).$$

- We obtained convergence results from fixed point theory, with the hypothesis "D_σ is averaged". (related to D_σ firmly nonexpansive, D_σ nonexpansive, or $\mathrm{Id} - D_\sigma$ nonexpansive)

Short Flash-back

- Plug-and-Play framework (Venkatakrishnan et al., 2013) (ADMM with non-deep denoisers)
- (Chan et al., 2016) PnP-ADMM with BM3D
  Convergence for stepsize $\tau \to 0$ and a “bounded" denoiser (i.e. $\|D_\sigma(x) - x\|^2 \leqslant C\sigma^2$)

Short Flash-back

- Plug-and-Play framework (Venkatakrishnan et al., 2013) (ADMM with non-deep denoisers)
- (Chan et al., 2016) PnP-ADMM with BM3D
  Convergence for stepsize $\tau \to 0$ and a “bounded" denoiser (i.e. $\|D_\sigma(x) - x\|^2 \leqslant C\sigma^2$)
- REgularization by Denoising (Romano et al., 2017) (GD/PGD/HQS with non-deep denoisers)
  Analysis with two hypotheses on D_σ: local homogeneity, and $\forall x, \|JD_\sigma(x)\| \leqslant 1$.
- Clarifications on RED (Reehorst and Schniter, 2018) (D_σ should have symmetric Jacobian)
- RED-PRO reformulates as a convex minimization problem on $\mathrm{Fix}(D_\sigma)$ (Cohen et al., 2021b)

Short Flash-back

- Plug-and-Play framework (Venkatakrishnan et al., 2013) (ADMM with non-deep denoisers)
- (Chan et al., 2016) PnP-ADMM with BM3D
  Convergence for stepsize $\tau \to 0$ and a “bounded" denoiser (i.e. $\|D_\sigma(x) - x\|^2 \leqslant C\sigma^2$)
- REgularization by Denoising (Romano et al., 2017) (GD/PGD/HQS with non-deep denoisers)
  Analysis with two hypotheses on D_σ: local homogeneity, and $\forall x, \|JD_\sigma(x)\| \leqslant 1$.
- Clarifications on RED (Reehorst and Schniter, 2018) (D_σ should have symmetric Jacobian)
- RED-PRO reformulates as a convex minimization problem on $\mathrm{Fix}(D_\sigma)$ (Cohen et al., 2021b)
- PnP (PGD/ADMM/DRS) with Lipschitz denoisers (Ryu et al., 2019)
  Convergence by contractive fixed point for with $\mathrm{Id} - D_\sigma$ ε-Lipschitz.

Short Flash-back

- Plug-and-Play framework (Venkatakrishnan et al., 2013) (ADMM with non-deep denoisers)
- (Chan et al., 2016) PnP-ADMM with BM3D
  Convergence for stepsize $\tau \to 0$ and a "bounded" denoiser (i.e. $\|D_\sigma(x) - x\|^2 \leqslant C\sigma^2$)
- REgularization by Denoising (Romano et al., 2017) (GD/PGD/HQS with non-deep denoisers)
  Analysis with two hypotheses on D_σ: local homogeneity, and $\forall x, \|JD_\sigma(x)\| \leqslant 1$.
- Clarifications on RED (Reehorst and Schniter, 2018) (D_σ should have symmetric Jacobian)
- RED-PRO reformulates as a convex minimization problem on $\mathrm{Fix}(D_\sigma)$ (Cohen et al., 2021b)
- PnP (PGD/ADMM/DRS) with Lipschitz denoisers (Ryu et al., 2019)
  Convergence by contractive fixed point for with $\mathrm{Id} - D_\sigma$ ε-Lipschitz.
- Convergent PnP with true MMSE denoiser (Xu et al., 2020) (MMSE is a non-convex prox)
- Convergence for firmly nonexpansive D_σ (Sun et al., 2021), (Pesquet et al., 2021)

Short Flash-back

- Plug-and-Play framework (Venkatakrishnan et al., 2013) (ADMM with non-deep denoisers)
- (Chan et al., 2016) PnP-ADMM with BM3D
  Convergence for stepsize $\tau \to 0$ and a "bounded" denoiser (i.e. $\|D_\sigma(x) - x\|^2 \leqslant C\sigma^2$)
- REgularization by Denoising (Romano et al., 2017) (GD/PGD/HQS with non-deep denoisers)
  Analysis with two hypotheses on D_σ: local homogeneity, and $\forall x, \|JD_\sigma(x)\| \leqslant 1$.
- Clarifications on RED (Reehorst and Schniter, 2018) (D_σ should have symmetric Jacobian)
- RED-PRO reformulates as a convex minimization problem on $\mathrm{Fix}(D_\sigma)$ (Cohen et al., 2021b)
- PnP (PGD/ADMM/DRS) with Lipschitz denoisers (Ryu et al., 2019)
  Convergence by contractive fixed point for with $\mathrm{Id} - D_\sigma$ ε-Lipschitz.
- Convergent PnP with true MMSE denoiser (Xu et al., 2020) (MMSE is a non-convex prox)
- Convergence for firmly nonexpansive D_σ (Sun et al., 2021), (Pesquet et al., 2021)
- Convergence for gradient-step denoiser (Hurault et al., 2021), (Cohen et al., 2021a)
- and several other contributions... see the review (Kamilov et al., 2023)

Denoising prior

Find x from observation $y = x + \xi$

- Input distribution $p(x)$.
- Gaussian noise $\xi \sim \mathcal{N}(0, \sigma^2 \mathrm{Id})$.
- Noisy observation y with density $p_\sigma(y)$ where $p_\sigma = p * \mathcal{N}(0, \sigma^2 \mathrm{Id})$.

MAP estimator

$$D_\sigma^{\mathrm{MAP}}(y) = \operatorname*{Argmax}_x p(x|y)$$

MMSE estimator

$$D_\sigma^{\mathrm{MMSE}}(y) = \mathbb{E}_{x \sim p(x|y)}[x]$$

Bayes:

$$D_\sigma^{\mathrm{MAP}} = \mathrm{Prox}_{-\sigma^2 \log p}$$

Tweedie:

$$D_\sigma^{\mathrm{MMSE}} = \mathrm{Id} - \nabla(-\sigma^2 \log p_\sigma)$$

A denoiser is related to an **implicit prior**

PnP and RED algorithms

Find $x^* \in \operatorname{Argmin} f(x) + \lambda g(x)$ with $f = -\log p(y|.)$ and $g \propto -\log p$

$$\begin{cases} \text{GD} & : x_{k+1} = (\text{Id} - \tau(\nabla f + \lambda \nabla g))(x_k) \\ \text{HQS} & : x_{k+1} = \text{Prox}_{\tau\lambda g} \circ \text{Prox}_{\tau f}(x_k) \\ \text{PGD} & : x_{k+1} = \text{Prox}_{\tau\lambda g} \circ (\text{Id} - \tau \nabla f)(x_k) \\ \text{DRS} & : x_{k+1} = \frac{1}{2}\text{Id} + \frac{1}{2}(2\,\text{Prox}_{\tau\lambda g} - \text{Id}) \circ (2\,\text{Prox}_{\tau f} - \text{Id})(x_k) \end{cases}$$

MAP denoiser

$$D_\sigma(y) = \text{Prox}_{-\sigma^2 \log p}(y)$$

MMSE denoiser

$$D_\sigma(y) = (\text{Id} + \sigma^2 \nabla \log p_\sigma)(y)$$

Plug-and-play (PnP)

(Venkatakrishnan et al., 2013)

$$\text{Prox}_{\tau\lambda g} \longrightarrow D_\sigma$$

Regularization by denoising (RED)

(Romano et al., 2017)

$$\text{Id} - \nabla g \longrightarrow D_\sigma$$

Today

- ⚠ In practice, *a priori*, there is no g_σ such that $D_\sigma = \mathrm{Prox}_{\tau g_\sigma}$ or $D_\sigma = \mathrm{Id} - \nabla g_\sigma$...
- We will construct a deep denoiser for which there actually *is* such an explicit g_σ.
- We will formulate convergence results with **non-convex optimization** on $f + \lambda g_\sigma$.
- We will see when D_σ can be formulated as a non-convex prox.
- We will discuss training, and connections with score-matching.

Plan

Convergence by Non-Convex Optimization

Gradient-Step and Proximal Denoisers

Further Topics

Proximal Gradient Descent (non-convex case)

We say that $f : \mathbf{R}^n \to \mathbf{R} \cup \{+\infty\}$ is proper if f is not $+\infty$ everywhere.

Definition

For $f : \mathbf{R}^n \to \mathbf{R} \cup \{+\infty\}$, for any $x \in \mathbf{R}^n$, we define the point-to-set map Prox_f by

$$\mathrm{Prox}_f(x) = \operatorname*{Argmin}_{z \in \mathbf{R}^n} f(z) + \frac{1}{2}\|z - x\|^2.$$

One can see that $\mathrm{Prox}_f(x)$ is non-empty as soon as f is l.s.c. with $\inf f > -\infty$ and f proper.

Let $f, g : \mathbf{R}^n \to \mathbf{R} \cup \{+\infty\}$ be proper l.s.c. and lower-bounded. Suppose that f is differentiable.

In order to minimize $F = f + g$, we consider the **proximal gradient descent** (PGD) algorithm

$$x_{k+1} \in \mathrm{Prox}_{\tau g}(x_k - \tau \nabla f(x_k)),$$

where $\tau > 0$ is a step size.

Convergence of PGD for non-convex functions

Theorem (e.g. Attouch et al. 2013)

Let $f, g : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be proper, l.s.c., bounded from below. Let $F = f + g$.
Assume that f is differentiable with ∇f being L_f-Lipschitz. Suppose $\tau < \frac{1}{L_f}$.
Then the PGD sequence $x_{k+1} \in \mathrm{Prox}_{\tau g}(x_k - \tau \nabla f(x_k))$, satisfies

1. *$F(x_k)$ is non-increasing, and thus converges.*
2. $\sum_{k=0}^{\infty} \|x_{k+1} - x_k\|^2 < \infty$ *and* $\min_{k<K} \|x_{k+1} - x_k\| = \mathcal{O}(\frac{1}{\sqrt{K}})$.
3. *The cluster points of (x_k) are critical points of F.*

Convergence Proof of PGD

In order to prove (i), we need to prove a descent inequality. The definition of x_{k+1} gives

$$x_{k+1} \in \operatorname*{Argmin}_x g(x) + \langle x - x_k, \nabla f(x_k)\rangle + \frac{1}{2\tau}\|x - x_k\|^2.$$

In particular,

$$f(x_k) + g(x_{k+1}) + \langle x_{k+1} - x_k, \nabla f(x_k)\rangle + \frac{1}{2\tau}\|x_{k+1} - x_k\|^2 \leqslant f(x_k) + g(x_k) = F(x_k).$$

Since f has L_f-Lipschitz gradient, the descent lemma gives

$$f(x_{k+1}) \leqslant f(x_k) + \langle x_{k+1} - x_k, \nabla f(x_k)\rangle + \frac{L_f}{2}\|x_{k+1} - x_k\|^2.$$

Therefore, $f(x_k) + \langle x_{k+1} - x_k, \nabla f(x_k)\rangle + \frac{1}{2\tau}\|x_{k+1} - x_k\|^2 \geqslant f(x_{k+1}) + \left(\frac{1}{2\tau} - \frac{L_f}{2}\right)\|x_{k+1} - x_k\|^2,$

and finally $F(x_{k+1}) + \left(\frac{1}{2\tau} - \frac{L_f}{2}\right)\|x_{k+1} - x_k\|^2 \leqslant F(x_k).$

Convergence Proof of PGD

The last inequality gives that for $\tau < \frac{1}{L}$, $F(x_k)$ is non-increasing.
Since $\inf F > -\infty$, we obtain $F(x_k) \to \ell \in \mathbf{R}$.
Summing the previous inequality also gives that

$$\left(\frac{1}{2\tau} - \frac{L_f}{2}\right) \sum_{k=0}^{K-1} \|x_{k+1} - x_k\|^2 \leqslant \sum_{k=0}^{K-1} F(x_k) - F(x_{k+1}) = F(x_0) - F(x_K) \leqslant F(x_0) - \ell.$$

It gives that $C = \sum_{k=0}^{\infty} \|x_{k+1} - x_k\|^2 < \infty$, and also,

$$\min_{k<K} \|x_{k+1} - x_k\|^2 \leqslant \frac{1}{K} \sum_{k=0}^{K-1} \|x_{k+1} - x_k\|^2 \leqslant \frac{C}{K}.$$

Limiting Subdifferential

The last part relies on a notion of critical point for non-convex functions (Attouch et al., 2013).

For $f : \mathbf{R}^n \to \mathbf{R} \cup \{+\infty\}$ and $x \in \mathbf{R}^n$ such that $f(x) < \infty$, we define the Fréchet subdifferential by

$$\hat{\partial} f(x) = \left\{ v \in \mathbf{R}^n \mid \quad \liminf_{y \to x} \frac{f(y) - f(x) - \langle v, y - x \rangle}{\|x - y\|} \geqslant 0 \right\},$$

and the limiting subdifferential as

$$\partial^{\lim} f(x) = \{ v \in \mathbf{R}^n \mid \exists (x_k), (v_k), v_k \in \hat{\partial} f(x_k), x_k \to x, f(x_k) \to f(x) \}.$$

Then one can show that (see (Rockafellar and Wets, 2009))

- $\partial f(x) \subset \hat{\partial} f(x) \subset \partial^{lim} f(x)$
- if x is a local minimizer of f, then $0 \in \hat{\partial} f(x)$
- if f is $\mathscr{C}^1$ and $g(x) < +\infty$, then $\partial^{\lim}(f + g)(x) = \nabla f(x) + \partial^{lim} g(x)$.
- if $x_k \to x$, $v_k \to v$, $v_k \in \partial^{lim} f(x_k)$, and $f(x_k) \to f(x)$, then $v \in \partial^{lim} f(x)$.
- if f is proper l.s.c., $z \in \mathrm{Prox}_f(x) \;\Rightarrow\; x - z \in \partial^{lim} f(z)$.

Convergence Proof of PGD

We can now end the proof. Thanks to the critical analysis of $\mathrm{Prox}_{\tau g}$, we have

$$v_k := \frac{x_k - x_{k-1}}{\tau} - \nabla f(x_k) \in \partial^{\lim} g(x_k).$$

If (x_{k_i}) is a subsequence that converges to a x, $\nabla f(x_{k_i}) \to \nabla f(x)$.
Since $\|x_{k+1} - x_k\| \to 0$, we deduce that $v_{k_i} \to -\nabla f(x)$.

Since g is l.s.c., we have $\liminf g(x_{k_i}) \geqslant g(x)$.
And again, with the optimality condition of x_{k+1},

$$g(x_{k_i}) + \langle x_{k_i} - x_{k_i-1}, \nabla f(x_{k_i-1})\rangle + \frac{1}{2\tau}\|x_{k_i} - x_{k_i-1}\|^2 \leqslant g(x) + \langle x - x_{k_i-1}, \nabla f(x_{k_i-1})\rangle + \frac{1}{2\tau}\|x - x_{k_i-1}\|^2$$

Since x_{k_i-1} also tends to x, we get $\limsup g(x_{k_i}) \leqslant g(x)$, and thus $g(x_{k_i}) \to g(x)$.

By the fact that "$\partial^{\lim} g(x)$ is closed", we get $-\nabla f(x) \in \partial^{lim} g(x)$, and thus $0 \in \partial^{lim} F(x)$.

The Kurdyka-Łojasiewicz property (Attouch et al., 2010)

In order to get convergence of the iterates, we need a technical assumption.

Definition (Kurdyka-Łojasiewicz (KŁ) property)

(a) A function $f : \mathbb{R}^n \longrightarrow \mathbb{R} \cup \{+\infty\}$ is said to have the Kurdyka-Lojasiewicz property at $x^* \in dom(f)$ if there exists $\eta \in (0, +\infty)$, a neighborhood U of x^* and a continuous concave function $\psi : [0, \eta) \longrightarrow \mathbb{R}_+$ such that $\psi(0) = 0$, ψ is $\mathcal{C}^1$ with $\psi' > 0$ on $(0, \eta)$, and $\forall x \in U \cap [f(x^*) < f < f(x^*) + \eta]$, the Kurdyka-Łojasiewicz inequality holds:

$$\psi'(f(x) - f(x^*))\text{dist}(0, \partial^{lim} f(x)) \geq 1.$$

(b) Proper lower semicontinuous functions which satisfy the Kurdyka-Lojasiewicz inequality at each point of $dom(\partial^{lim} f)$ are called KŁ functions.

Theorem (Attouch et al. 2013)

Let $f, g : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be proper, l.s.c., lower bounded. Let $F = f + g$.
Assume that f is differentiable with ∇f being L_f-Lipschitz and that F is KŁ. Assume $\tau < \frac{1}{L_f}$.
Suppose also that the PGD iterates (x_k) are bounded.
Then (x_k) converges to a critical point of F.

Summary: Convergence of nonconvex proximal splitting algorithms

The same kind of techniques applies to the Douglas-Rachford algorithm. Recall the algorithms:

- **PGD** : $x_{k+1} = \text{Prox}_{\tau f} \circ (\text{Id} - \tau \nabla g)$
- **DRS** : $x_{k+1} = \frac{1}{2}\text{Id} + \frac{1}{2}(2\,\text{Prox}_{\tau f} - \text{Id}) \circ (2\,\text{Prox}_{\tau g} - \text{Id})$

Objective : Show that $x_k \to x^* \in \{x \in \mathbf{R}^n \mid 0 \in \partial f(x) + \nabla g(x)\}$.

Suppose

- g is L-smooth
- $f + g$ coercive and bounded from below.
- f and g verify the Kurdyka-Lojasiewicz (KL) property.

Then, for $\begin{cases} (\textbf{PGD}) & \tau L < 1 \\ (\textbf{DRS}) & \tau L < 1/2 \end{cases}$ (Attouch et al., 2013) / (Themelis and Patrinos, 2020)

if (x_k) is bounded, it converges towards a critical point of $f + g$.

Backtracking

What if we don't know the Lipschitz constant at stake?
For example, we have shown that the PGD update $T_\tau(x_k)$ satisfies a descent lemma

$$\forall \tau < \frac{1}{L_f}, \quad F(x_k) - F(T_\tau(x_k)) \geqslant \left(\frac{1}{2\tau} - \frac{L_f}{2}\right) \|T_\tau(x_k) - x_k\|^2.$$

For parameters $\gamma \in (0, \frac{1}{2}), \eta \in [0, 1)$, the backtracking procedure consists in

$$\textbf{while} \quad F(x_k) - F(T_\tau(x_k)) < \frac{\gamma}{\tau} \|T_\tau(x_k) - x_k\|^2 \quad \textbf{do} \quad \tau \leftarrow \eta\tau.$$

Since this last inequality is not true for $\tau < \frac{1-2\gamma}{L}$, the backtracking loop stops in finite time.

It is possible to show that the convergence guarantees still hold with backtracking.

Plan

Convergence by Non-Convex Optimization

Gradient-Step and Proximal Denoisers

Further Topics

The Gradient-Step Denoiser

- (Romano et al., 2017) If D_σ has **symmetric Jacobian**, then

$$D_\sigma = \mathrm{Id} - \nabla g_\sigma \quad \text{with} \quad g_\sigma(x) = \frac{1}{2}\langle x, x - D_\sigma(x)\rangle$$

✗ Not verified by common denoisers (Reehorst and Schniter, 2018).

- (Hurault et al., 2021), (Cohen et al., 2021a) **“Gradient Step" (GS)** Denoiser:

$$D_\sigma = \mathrm{Id} - \nabla g_\sigma \quad \text{with} \quad g_\sigma(x) = \frac{1}{2}\|x - N_\sigma(x)\|^2$$

with $N_\sigma : \mathbb{R}^n \to \mathbb{R}^n$ is a $\mathcal{C}^1$ neural network (smoothed DRUNet (Zhang et al., 2021))

- The denoiser can be written

$$D_\sigma(x) = N_\sigma(x) + J_{N_\sigma}(x)^T(x - N_\sigma(x)).$$

- A composition of functions with bounded Lipschitz differentials has Lipschitz differential.
- g_σ satisfies the KŁ property (as soon as activations are subanalytic).

Connection with Score-Matching

- The denoiser is trained on a data distribution p_X of clean images by

$$\underset{\text{Param}(D_\sigma)}{\text{Argmin}} \ \mathbb{E}_{x\sim p_X, \xi \sim \mathcal{N}(0,\sigma^2 \text{Id})} \left[\|D_\sigma(x+\xi) - x\|^2 \right].$$

- This is actually equivalent to

$$\underset{\text{Param}(D_\sigma)}{\text{Argmin}} \ \mathbb{E}_{y\sim p_\sigma} \left[\|D_\sigma(y) - D_\sigma^{\text{MMSE}}(y)\|^2 \right]$$

or, thanks to Tweedie's formula, to

$$\underset{\text{Param}(D_\sigma)}{\text{Argmin}} \ \mathbb{E}_{y\sim p_\sigma} \left[\|D_\sigma(y) - y - \sigma^2 \nabla \log p_\sigma(y)\|^2 \right]$$

$$\text{i.e.} \quad \underset{\text{Param}(D_\sigma)}{\text{Argmin}} \ \mathbb{E}_{y\sim p_\sigma} \left[\|\nabla g_\sigma(y) + \sigma^2 \nabla \log p_\sigma(y)\|^2 \right]$$

- Therefore, $\sigma^{-2} \nabla g_\sigma$ is designed to approximate the score $-\nabla \log p_\sigma$.

Convergence of GS-PnP (Hurault et al., 2021)

Let $\lambda > 0$. We here target minima of $F = f + \lambda g_\sigma$.
For $\tau > 0$, consider

$$x_{k+1} = \text{Prox}_{\tau f} \circ (\tau \lambda D_\sigma + (1 - \tau\lambda)\text{Id})(x_k)$$

with gradient-step denoiser $D_\sigma = \text{Id} - \nabla g_\sigma$.

Theorem

Let $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ and $g_\sigma : \mathbb{R}^n \to \mathbb{R}$ be proper lower semicontinous functions.
Let $\lambda > 0$, $F = f + \lambda g_\sigma$. Suppose that

- *g_σ is differentiable with L-Lipschitz gradient,*
- *F is bounded from below and satisfies the Kurdyka-Łojasiewicz property.*

Then, for $\tau < \frac{1}{\lambda L}$,

- *$(F(x_k))$ is non-increasing and converges,*
- *If (x_k) is bounded, then it converges to a critical point of F.*

Remark: It is possible to modify the regularization g_σ to ensure that $\lim_{\|x\| \to \infty} F(x) = +\infty$.

Characterization of Proximal Operators

- (Moreau, 1965)
  If $D_\sigma = \partial h_\sigma$ with h_σ convex and D_σ is nonexpansive,
  then $\exists \phi_\sigma : \mathbb{R}^n \to \mathbb{R}$ convex such that $D_\sigma = \text{Prox}_{\phi_\sigma}$.
  ✗ Hard too enforce both conditions at the same time
- (Gribonval and Nikolova, 2020)
  If $D_\sigma = \partial h_\sigma$ with h_σ convex ~~and D_σ is nonexpansive~~,
  then $\exists \phi_\sigma : \mathbb{R}^n \to \mathbb{R}$ ~~convex~~ such that $D_\sigma = \text{Prox}_{\phi_\sigma}$.

→ **Proximal denoiser** (Hurault et al., 2022)

$$D_\sigma = \text{Id} - \nabla g_\sigma = \nabla h_\sigma \quad \text{with} \quad h_\sigma(x) = \frac{\|x\|^2}{2} - g_\sigma(x)$$

∇g_σ L-Lipschitz with $L < 1 \Rightarrow \exists \phi_\sigma$ $\frac{L}{L+1}$-weakly convex s.t. $D_\sigma = \text{Prox}_{\phi_\sigma}$

✗ $D_\sigma = \text{Prox}_{\phi_\sigma}$ restricts the stepsize $\tau = 1$.

Gradient-Step and Proximal Denoisers

Theorem (Hurault et al. 2022)

Let $g_\sigma : \mathbb{R}^n \to \mathbb{R}$ *a* $\mathcal{C}^{k+1}$ *function with* $k \geq 1$ *and* ∇g_σ *L-Lipschitz with* $L < 1$*. Let*

$$D_\sigma = \mathrm{Id} - \nabla g_\sigma = \nabla h_\sigma \quad \text{with} \quad h_\sigma(x) = \frac{\|x\|^2}{2} - g_\sigma(x)$$

Then

(i) h_σ *is* $(1-L)$*-strongly convex and* $\forall x \in \mathbb{R}^n$, $J_{D_\sigma}(x)$ *is positive definite*

(ii) D_σ *is injective,* $\mathrm{Im}(D_\sigma)$ *is open and,* $\forall x \in \mathbb{R}^n$, $D_\sigma(x) = \mathrm{Prox}_{\phi_\sigma}(x)$*, with*

$$\phi_\sigma(x) \propto \begin{cases} g_\sigma(D_\sigma{}^{-1}(x)) - \frac{1}{2}\|D_\sigma{}^{-1}(x) - x\|^2 & \text{if } x \in \mathrm{Im}(D_\sigma), \\ +\infty & \text{otherwise}, \end{cases} \tag{1}$$

(iii) ϕ_σ *is* $\frac{L}{L+1}$ *weakly convex (i.e.* $\phi_\sigma + \frac{L}{L+1}\frac{\|\cdot\|^2}{2}$ *is convex).*

Training the Gradient-step and Proximal denoisers

Training loss: GS-Denoiser - Prox-Denoiser

$$\underset{\text{Param}(D_\sigma)}{\text{Argmin}}\ \mathbb{E}_{x,\xi_\sigma}\left[\|D_\sigma(x+\xi_\sigma)-x\|^2+\mu\max(\|\nabla^2 g_\sigma(x+\xi_\sigma)\|_S,1-\epsilon)\right]$$

$\sigma(./255)$	5	15	25
DRUNet	40.19	33.89	31.25
GS-Denoiser	40.26	33.90	31.26
Prox-Denoiser	40.12	33.60	30.82

Table: Denoising PSNR on the CBSD68 dataset

$\sigma(./255)$	5	15	25
GS-DRUNet	1.26	1.96	3.27
Prox-DRUNet	0.92	0.99	0.96

Table: $\max_x \|\nabla^2 g_\sigma(x)\|_S$ on the CBSD68 dataset

Convergence of Prox-PNP-PGD (Hurault et al., 2022)

Let $\lambda > 0$. We here target minima of $F = f + \lambda\phi_\sigma$.

$$\begin{cases} z_{k+1} = x_k - \frac{1}{\lambda}\nabla f(x_k) \\ x_{k+1} = D_\sigma(z_{k+1}) \end{cases} \quad \text{with} \quad D_\sigma = \mathrm{Id} - \nabla g_\sigma = \mathrm{Prox}_{\phi_\sigma}.$$

For $\lambda > 0$, let $F = f + \lambda\phi_\sigma$.

Theorem

Let $f, g_\sigma : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be proper lower semicontinous functions, bounded from below. Suppose that

- *f is differentiable with L_f-Lipschitz gradient*
- *g_σ is $\mathcal{C}^2$ with L-Lipschitz gradient and $L < 1$,*
- *F is bounded from below and satisfies the Kurdyka-Łojasiewicz property.*

Then, for $\lambda > L$,

- *$(F(x_k))$ is non-increasing and converges,*
- *If (x_k) is bounded, it converges to a critical point of F.*

Convergence of Prox-PNP-DRS1 (Hurault et al., 2022)

Let $\lambda > 0$. We here target minima of $F = f + \lambda\phi_\sigma$.

$$\begin{cases} y_{k+1} = \mathrm{Prox}_{\frac{1}{\lambda}f}(x_k) \\ z_{k+1} = D_\sigma(2y_{k+1} - x_k) \\ x_{k+1} = x_k + (z_{k+1} - y_{k+1}) \end{cases} \quad \text{with} \quad D_\sigma = \mathrm{Id} - \nabla g_\sigma = \mathrm{Prox}_{\phi_\sigma}.$$

Let $\lambda > 0$, and $F^{DR,1}_{\lambda,\sigma}(x_{k-1}) = \phi_\sigma(z_k) + \frac{1}{\lambda}f(y_k) + \langle y_k - x_{k-1}, y_k - z_k\rangle + \frac{1}{2}\|y_k - z_k\|^2$.

Theorem

Let $f, g_\sigma : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be proper lower semicontinous functions, bounded from below. Suppose that

- *f is convex, differentiable with L_f-Lipschitz gradient*
- *g_σ is $\mathcal{C}^2$ with L-Lipschitz gradient and $L < 1$,*
- *F is bounded from below and satisfies the Kurdyka-Łojasiewicz property.*

Then, for $\lambda > L$,

- *$(F^{DR,1}_{\lambda,\sigma}(x_k))$ is non-increasing and converges,*
- *If (x_k) is bounded, it converges to a critical point of F.*

Convergence of Prox-PNP-DRS2 (Hurault et al., 2022)

Let $\lambda > 0$. We here target minima of $F = f + \lambda\phi_\sigma$.

$$\begin{cases} y_{k+1} = D_\sigma(x_k) \\ z_{k+1} = \mathrm{Prox}_{\frac{1}{\lambda}f}(2y_{k+1} - x_k) \\ x_{k+1} = x_k + (z_{k+1} - y_{k+1}) \end{cases} \quad \text{with} \quad D_\sigma = \mathrm{Id} - \nabla g_\sigma = \mathrm{Prox}_{\phi_\sigma}.$$

Let $\lambda > 0$, and $F^{DR,2}_{\lambda,\sigma}(x_{k-1}) = \phi_\sigma(y_k) + \frac{1}{\lambda}f(z_k) + \langle y_k - x_{k-1}, y_k - z_k\rangle + \frac{1}{2}\|y_k - z_k\|^2$.

Theorem

Let $f, g_\sigma : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ be proper lower semicontinous functions, bounded from below. Suppose that

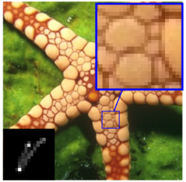
- $\mathrm{Im}(D_\sigma)$ *is convex*
- g_σ *is* $\mathcal{C}^2$ *with L-Lipschitz gradient and* $L < \frac{1}{2}$
- *F is bounded from below and satisfies the Kurdyka-Łojasiewicz property.*

Then, for any $\lambda > 0$,

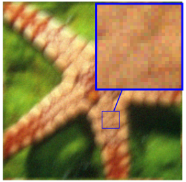
- $(F^{DR,2}_{\lambda,\sigma}(x_k))$ *is non-increasing and converges,*
- *If* (x_k) *is bounded, it converges to a critical point of F.*

Deblurring Example (Hurault et al., 2022)

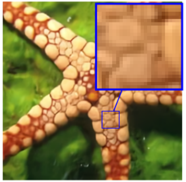
Deblurring with motion kernel and Gaussian noise std $\nu = 0.03$



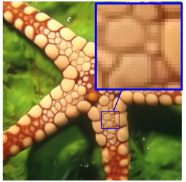
(a) Clean



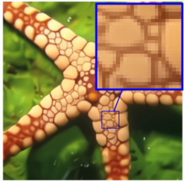
(b) Observed



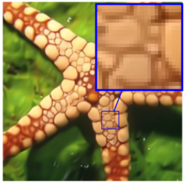
(c) IRCNN (28.66dB)



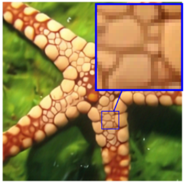
(d) DPIR (29.76dB)



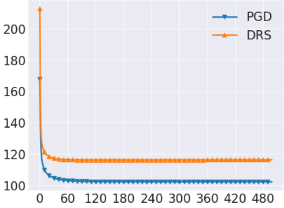
(e) GSPnP-HQS (29.90dB)



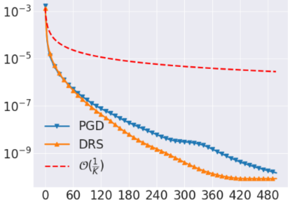
(f) Prox-PnP-PGD (29.41dB)



(g) Prox-PnP-DRS (29.65dB)



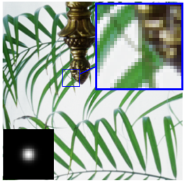
(h) $F_{\lambda,\sigma}(x_k)$

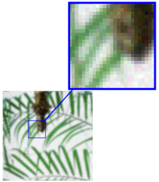


(i) $\min_{i \le k} ||x_{i+1} - x_i||^2$

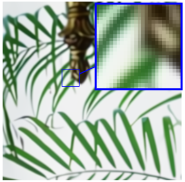
Super-resolution Example (Hurault et al., 2022)

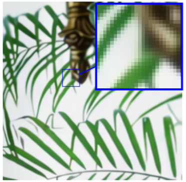
Super-resolution with scale 2, Gaussian blur kernel and Gaussian noise std $\nu = 0.01$

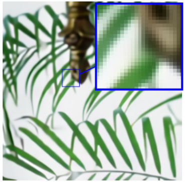

(a) Clean

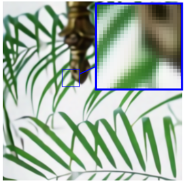

(b) Observed

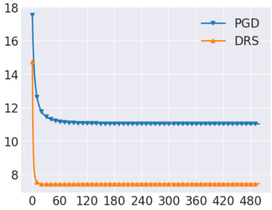

(c) IRCNN (22.82dB)


(d) DPIR (23.97dB)

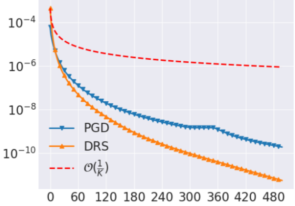

(e) GSPnP-HQS (24.81dB)


(f) Prox-PnP-PGD (23.96dB)


(g) Prox-PnP-DRS (24.36dB)



(h) $F_{\lambda,\sigma(x_k)}$



(i) $\min_{i \le k} ||x_{i+1} - x_i||^2$

Plan

Convergence by Non-Convex Optimization

Gradient-Step and Proximal Denoisers

Further Topics

Further Topics

- Exploit semi-convexity of ϕ_σ (Hurault et al., 2023)
- Unrolling plug-and-play algorithms (Repetti et al., 2022)
- Plug-and-Play posterior sampling (Laumont et al., 2022)
- Plug-and-Play adapted to more complex data-fidelity (Laroche et al., 2023)
- Stochastic plug-and-play regularizations (Renaud et al., 2024)

Take-home Messages

- Using appropriate denoisers can make Plug-and-Play algorithms more stable.
- Gradient-step Denoisers allow to recover an explicit minimization problem.
- This helps to recover precise numerical control, and improves stability.
- With backtracking, we don't even have to know the Lipschitz constant of the regularization.
- However, in practice, parameters should be adjusted to avoid bad local minima.
- Visual results can be further improved by tuning the strategy on σ ($\rightarrow$ diffusion models)

Et avant le TP, une petite page de publicité...

- S. Hurault's thesis on PnP algorithms: `https://www.theses.fr/2023BORD0336`
- A nice document on gradient descent by Robert Gower:
  `https://perso.telecom-paristech.fr/rgower/pdf/M2_statistique_optimisation/grad_conv.pdf`
  See also the handbook (Garrigos and Gower, 2023) or C. Dossal's lecture notes.
- Imaging in Paris seminar: `https://imaging-in-paris.github.io/`
- M2 internship on PnP methods for Hyperspectral Unmixing (with C. Kervazo and yours truly)
- Python/Pytorch library for Plug-and-Play Imaging:

Deep Inverse

`https://deepinv.github.io/`

Main contributors: S. Hurault, J. Tachella, M. Terris

THANK YOU FOR YOUR ATTENTION!

References I

Attouch, H., Bolte, J., Redont, P., and Soubeyran, A. (2010). Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the kurdyka-łojasiewicz inequality. *Mathematics of operations research*, 35(2):438–457.

Attouch, H., Bolte, J., and Svaiter, B. F. (2013). Convergence of descent methods for semi-algebraic and tame problems: proximal algorithms, forward–backward splitting, and regularized gauss–seidel methods. *Mathematical Programming*, 137(1-2):91–129.

Chan, S. H., Wang, X., and Elgendy, O. A. (2016). Plug-and-play ADMM for image restoration: Fixed-point convergence and applications. *IEEE Transactions on Computational Imaging*, 3(1):84–98.

Cohen, R., Blau, Y., Freedman, D., and Rivlin, E. (2021a). It has potential: Gradient-driven denoisers for convergent solutions to inverse problems. *Advances in Neural Information Processing Systems*, 34.

Cohen, R., Elad, M., and Milanfar, P. (2021b). Regularization by denoising via fixed-point projection (RED-PRO). *SIAM Journal on Imaging Sciences*, 14(3):1374–1406.

Garrigos, G. and Gower, R. M. (2023). Handbook of convergence theorems for (stochastic) gradient methods.

References II

Gribonval, R. and Nikolova, M. (2020). A characterization of proximity operators. *Journal of Mathematical Imaging and Vision*, 62(6):773–789.

Hurault, S., Chambolle, A., Leclaire, A., and Papadakis, N. (2023). A relaxed proximal gradient descent algorithm for convergent plug-and-play with proximal denoiser. In *International Conference on Scale Space and Variational Methods in Computer Vision*, pages 379–392. Springer.

Hurault, S., Leclaire, A., and Papadakis, N. (2021). Gradient step denoiser for convergent plug-and-play. In *International Conference on Learning Representations*.

Hurault, S., Leclaire, A., and Papadakis, N. (2022). Proximal denoiser for convergent plug-and-play optimization with nonconvex regularization. In *International Conference on Machine Learning*, pages 9483–9505. PMLR.

Kamilov, U. S., Bouman, C. A., Buzzard, G. T., and Wohlberg, B. (2023). Plug-and-play methods for integrating physical and learned models in computational imaging: Theory, algorithms, and applications. *IEEE Signal Processing Magazine*, 40(1):85–97.

Laroche, C., Almansa, A., Coupeté, E., and Tassano, M. (2023). Provably convergent plug & play linearized admm, applied to deblurring spatially varying kernels. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5. IEEE.

References III

Laumont, R., Bortoli, V. D., Almansa, A., Delon, J., Durmus, A., and Pereyra, M. (2022). Bayesian imaging using plug & play priors: when langevin meets tweedie. *SIAM Journal on Imaging Sciences*, 15(2):701–737.

Moreau, J.-J. (1965). Proximité et dualité dans un espace hilbertien. *Bulletin de la Société Mathématique de France*, 93:273–299.

Pesquet, J.-C., Repetti, A., Terris, M., and Wiaux, Y. (2021). Learning maximally monotone operators for image recovery. *SIAM Journal on Imaging Sciences*, 14(3):1206–1237.

Reehorst, E. T. and Schniter, P. (2018). Regularization by denoising: Clarifications and new interpretations. *IEEE transactions on computational imaging*, 5(1):52–67.

Renaud, M., Prost, J., Leclaire, A., and Papadakis, N. (2024). Plug-and-play image restoration with stochastic denoising regularization.

Repetti, A., Terris, M., Wiaux, Y., and Pesquet, J.-C. (2022). Dual forward-backward unfolded network for flexible plug-and-play. In *2022 30th European Signal Processing Conference (EUSIPCO)*, pages 957–961. IEEE.

Rockafellar, R. T. and Wets, R. J.-B. (2009). *Variational analysis*, volume 317. Springer Science & Business Media.

References IV

Romano, Y., Elad, M., and Milanfar, P. (2017). The little engine that could: Regularization by denoising (RED). *SIAM Journal on Imaging Sciences*, 10(4):1804–1844.

Ryu, E., Liu, J., Wang, S., Chen, X., Wang, Z., and Yin, W. (2019). Plug-and-play methods provably converge with properly trained denoisers. In *International Conference on Machine Learning*, pages 5546–5557. PMLR.

Sun, Y., Wu, Z., Xu, X., Wohlberg, B., and Kamilov, U. S. (2021). Scalable plug-and-play admm with convergence guarantees. *IEEE Transactions on Computational Imaging*, 7:849–863.

Themelis, A. and Patrinos, P. (2020). Douglas–rachford splitting and ADMM for nonconvex optimization: Tight convergence results. *SIAM Journal on Optimization*, 30(1):149–181.

Venkatakrishnan, S. V., Bouman, C. A., and Wohlberg, B. (2013). Plug-and-play priors for model based reconstruction. In *2013 IEEE Global Conference on Signal and Information Processing*, pages 945–948. IEEE.

Xu, X., Sun, Y., Liu, J., Wohlberg, B., and Kamilov, U. S. (2020). Provable convergence of plug-and-play priors with mmse denoisers. *arXiv preprint arXiv:2005.07685*.

Zhang, K., Li, Y., Zuo, W., Zhang, L., Van Gool, L., and Timofte, R. (2021). Plug-and-play image restoration with deep denoiser prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.