Proximal Denoiser for Convergent Plug-and-Play Optimization with Nonconvex Regularization

Samuel Hurault, Arthur Leclaire, Nicolas Papadakis

Institut de Mathématiques de Bordeaux

PAPER

CODE

Image Inverse Problems

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

- $y \in \mathbb{R}^m$ observation
- $x \in \mathbb{R}^n$ unknown input
- $A \in \mathbb{R}^{m \times n}$ degradation operator
- ξ random noise, generally $\xi \sim \mathcal{N}(0, \sigma^2 \operatorname{Id}_m)$

Image Inverse Problems

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

- $y \in \mathbb{R}^m$ observation
- $x \in \mathbb{R}^n$ unknown input
- $A \in \mathbb{R}^{m \times n}$ degradation operator
- ξ random noise, generally $\xi \sim \mathcal{N}(0, \sigma^2 \operatorname{Id}_m)$

Denoising:

Image Inverse Problems

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

- $y \in \mathbb{R}^m$ observation
- $x \in \mathbb{R}^n$ unknown input
- $A \in \mathbb{R}^{m \times n}$ degradation operator
- ξ random noise, generally $\xi \sim \mathcal{N}(0, \sigma^2 \operatorname{Id}_m)$

Deblurring:

Image Inverse Problems

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

- $y \in \mathbb{R}^m$ observation
- $x \in \mathbb{R}^n$ unknown input
- $A \in \mathbb{R}^{m\times n}$ degradation operator
- ξ random noise, generally $\xi \sim \mathcal{N}(0, \sigma^2 \, \mathrm{Id}_m)$

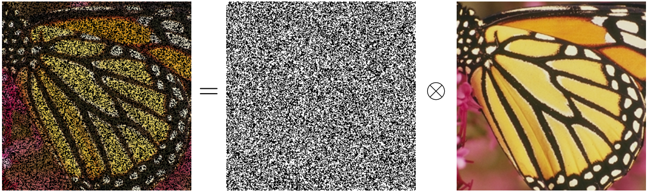
Super-resolution:

Image Inverse Problems

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

- $y \in \mathbb{R}^m$ observation
- $x \in \mathbb{R}^n$ unknown input
- $A \in \mathbb{R}^{m \times n}$ degradation operator
- ξ random noise, generally $\xi \sim \mathcal{N}(0, \sigma^2 \, \mathrm{Id}_m)$

Inpainting:



Maximum A-Posteriori

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

Maximum A-Posteriori

$$x^* \in \underset{x \in \mathbb{R}^n}{\arg\min}\, f(x) + g(x)$$

Maximum A-Posteriori

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

Maximum A-Posteriori

$$x^* \in \arg\min_{x \in \mathbb{R}^n} f(x) + g(x)$$

$\iff \quad \arg\min_{x \in \mathbb{R}^n}$ data-fidelity

e.g. $f(x) = \frac{1}{2}||Ax - y||^2$

Maximum A-Posteriori

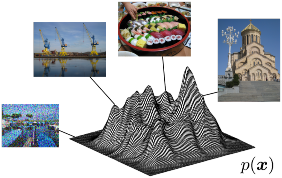
Find x from observation $y = Ax + \xi$

Maximum A-Posteriori

$$x^* \in \arg\min_{x \in \mathbb{R}^n} f(x) + g(x)$$

$\iff \arg\min_{x \in \mathbb{R}^n}$ data-fidelity $+$ prior

$e.g.\ f(x) = \frac{1}{2}||Ax - y||^2$



Maximum A-Posteriori

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

Maximum A-Posteriori

$$x^* \in \arg\min_{x \in \mathbb{R}^n} f(x) + g(x)$$

$\iff \arg\min_{x \in \mathbb{R}^n}$ data-fidelity + prior

e.g. $f(x) = \frac{1}{2}||Ax - y||^2$

Plug & Play (PnP) [Venkatakrishnan et al., ’13]

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

Maximum A-Posteriori

$$x^* \in \arg\min_{x \in \mathbb{R}^n} f(x) + g(x)$$

✗ unknown prior

Plug & Play (PnP) [Venkatakrishnan et al., '13]

Minimize

$$\arg\min_{x \in \mathbb{R}^n} f(x) + g(x)$$

Max

$x^* \in$

Plug & Play (PnP) [Venkatakrishnan et al., '13]

Max

$x^* \in$

Minimize

$$\arg\min_{x \in \mathbb{R}^n} f(x) + g(x)$$

Proximal algorithms

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

Plug & Play (PnP) [Venkatakrishnan et al., '13]

Minimize

$$\arg\min_{x \in \mathbb{R}^n} f(x) + g(x)$$

Proximal algorithms

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

Proximal operator :

$$\mathrm{Prox}_{\tau g}(y) = \arg\min_{x \in \mathbb{R}^n} \frac{1}{2\tau} ||x - y||^2 + g(x)$$

Max

$x^* \in$

Plug & Play (PnP) [Venkatakrishnan et al., '13]

Max

$x^* \in$

Minimize

$$\arg\min_{x \in \mathbb{R}^n} f(x) + g(x)$$

Proximal algorithms

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

Proximal operator :

$$\mathrm{Prox}_{\tau g}(y) = \arg\min_{x \in \mathbb{R}^n} \frac{1}{2\tau} ||x - y||^2 + g(x)$$

$\Leftrightarrow$ **Gaussian Denoising MAP** with prior g ($\sigma^2 = \tau$)

Plug & Play (PnP) [Venkatakrishnan et al., '13]

Max

$x^* \in$

Minimize

$$\arg\min_{x \in \mathbb{R}^n} f(x) + g(x)$$

Proximal algorithms

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

Proximal operator :

$$\mathrm{Prox}_{\tau g}(y) = \arg\min_{x \in \mathbb{R}^n} \frac{1}{2\tau} ||x - y||^2 + g(x) \approx D_\sigma(x)$$

$\Leftrightarrow$ **Gaussian Denoising MAP** with prior g $(\sigma^2 = \tau)$

PnP algorithms

- PnP-PGD: $x_{k+1} = D_\sigma \circ (\mathrm{Id} - \tau \nabla f)(x_k)$
- PnP-DRS: $x_{k+1} = \frac{1}{2} x_k + \frac{1}{2}(2 D_\sigma - \mathrm{Id}) \circ (2\,\mathrm{Prox}_{\tau f} - \mathrm{Id})(x_k)$

Plug & Play (PnP) [Venkatakrishnan et al., '13]

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

Maximum A-Posteriori

$$x^* \in \arg\min_{x \in \mathbb{R}^n} f(x) + g(x)$$

✗ unknown prior

$D_\sigma \approx \text{Prox}_{\tau g}$

Plug-and-Play (PnP)

✓ implicit prior

✓ $D_\sigma \approx \text{Prox}_{\tau g}$

Plug & Play (PnP) [Venkatakrishnan et al., '13]

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

Maximum A-Posteriori

$$x^* \in \arg\min_{x \in \mathbb{R}^n} f(x) + g(x)$$

✗ unknown prior

$D_\sigma \approx \mathrm{Prox}_{\tau g}$

Plug-and-Play (PnP)

✓ implicit prior
✓ SOTA restoration

✓ $D_\sigma \approx \mathrm{Prox}_{\tau g}$

Plug & Play (PnP) [Venkatakrishnan et al., '13]

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

Maximum A-Posteriori

$$x^* \in \arg\min_{x \in \mathbb{R}^n} f(x) + g(x)$$

✗ unknown prior

$D_\sigma \approx \mathrm{Prox}_{\tau g}$

Plug-and-Play (PnP)

✓ implicit prior
✓ SOTA restoration
✗ no minimization problem
✗ no convergence guarantees

✗ $D_\sigma \neq \mathrm{Prox}_{\tau g}$

Plug & Play (PnP) [Venkatakrishnan et al., '13]

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

Maximum A-Posteriori

$$x^* \in \arg\min_{x \in \mathbb{R}^n} f(x) + g(x)$$

✗ unknown prior

$D_\sigma \approx \mathrm{Prox}_{\tau g}$

Plug-and-Play (PnP)

✓ implicit prior
✓ SOTA restoration
✗ no minimization problem
✗ no convergence guarantees

How to ensure that D_σ is **exactly** a prox ?

Plug & Play (PnP) [Venkatakrishnan et al., '13]

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

Maximum A-Posteriori

$$x^* \in \arg\min_{x \in \mathbb{R}^n} f(x) + g(x)$$

✗ unknown prior

$D_\sigma \approx \mathrm{Prox}_{\tau g}$

Plug-and-Play (PnP)

✓ implicit prior
✓ SOTA restoration
✗ no minimization problem
✗ no convergence guarantees

How to ensure that D_σ is **exactly** a prox ?

✗ [Moreau '65] D_σ nonexpansive + gradient of convex function
✓ [Gribonval & Nikolava '20] D_σ gradient of convex function

Train the denoiser as a **gradient descent step**
[Hurault et al. '21, Cohen et al. '22] :

$$D_\sigma = \mathrm{Id} - \nabla g_\sigma$$

✗ no minimization problem
✗ no convergence guarantees

Train the denoiser as a **gradient descent step** [Hurault et al. '21, Cohen et al. '22] :

$$D_\sigma = \mathrm{Id} - \nabla g_\sigma$$

Proposition [Gribonval & Nikolova '20]

If $\mathrm{Id} - D_\sigma = \nabla g_\sigma$ is L-Lipschitz with $L < 1$ then there exists a closed-form function ϕ_σ, smooth on $Im(D_\sigma)$, such that

$$D_\sigma = \mathrm{Prox}_{\phi_\sigma}$$

✗ no minimization problem

✗ no convergence guarantees

Train the denoiser as a **gradient descent step** [Hurault et al. '21, Cohen et al. '22] :

$$D_\sigma = \mathrm{Id} - \nabla g_\sigma$$

Proposition [Gribonval & Nikolova '20]

If $\mathrm{Id} - D_\sigma = \nabla g_\sigma$ is L-Lipschitz with $L < 1$ then there exists a closed-form function ϕ_σ, smooth on $Im(D_\sigma)$, such that

$$D_\sigma = \mathrm{Prox}_{\phi_\sigma}$$

✗ no minimization problem

PnP algorithms

- PnP-PGD: $x_{k+1} = D_\sigma \circ (\mathrm{Id} - \tau \nabla f)(x_k)$
- PnP-DRS: $x_{k+1} = \frac{1}{2} x_k + \frac{1}{2}(2 D_\sigma - \mathrm{Id}) \circ (2\,\mathrm{Prox}_{\tau f} - \mathrm{Id})(x_k)$

Prox-PnP algorithms ($\tau = 1$)

- Prox-PnP-PGD: $x_{k+1} = \mathrm{Prox}_{\phi_\sigma} \circ (\mathrm{Id} - \nabla f)(x_k)$
- Prox-PnP-DRS: $x_{k+1} = \frac{1}{2} x_k + \frac{1}{2}(2\mathrm{Prox}_{\phi_\sigma} - \mathrm{Id}) \circ (2\,\mathrm{Prox}_{f} - \mathrm{Id})(x_k)$

Prox-PnP

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

Maximum A-Posteriori

$$x^* \in \arg\min_{x \in \mathbb{R}^n} f(x) + g(x)$$

✗ unknown prior

$D_\sigma \approx \text{Prox}_{\tau g}$

Plug-and-Play (PnP)

✓ implicit prior
✓ SOTA restoration
✗ no minimization problem
✗ no convergence guarantees

Prox-Denoiser
$D_\sigma = \text{Prox}_{\phi_\sigma}$

Prox-PnP [Hurault et al, '22]

$$x^* \in \arg\min_{x \in \mathbb{R}^n} f(x) + \phi_\sigma(x)$$

✓ explicit prior
✓ minimization problem

Prox-PnP

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

Proposed non-convex potential [Hurault et al. '21]:

$$g_\sigma(x) = \frac{1}{2}||x - N_\sigma(x)||^2$$

with a $\mathcal{C}^2$ neural network $N_\sigma : \mathbb{R}^n \to \mathbb{R}^n$ (*e.g.* DRUNet [Zhang et al. '21] with softplus activations).

✓ explicit prior
✓ minimization problem

$D_\sigma = \mathrm{Prox}_{\phi_\sigma}$

Prox-PnP

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

Proposed non-convex potential [Hurault et al. '21]:

$$g_\sigma(x) = \frac{1}{2}||x - N_\sigma(x)||^2$$

with a $\mathcal{C}^2$ neural network $N_\sigma : \mathbb{R}^n \to \mathbb{R}^n$ (*e.g.* DRUNet [Zhang et al. '21] with softplux activations).

Training loss :

$$\mathbb{E}_{x,\xi_\sigma}\left[||D_\sigma(x+\xi_\sigma) - x||^2 + \mu \max(||J_{(\mathrm{Id}-D_\sigma)}(x+\xi_\sigma)||_S, 1-\epsilon)\right]$$

M

Prox

m

ees

✓ explicit prior

✓ minimization problem

$D_\sigma = \mathrm{Prox}_{\phi_\sigma}$

Prox-PnP

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

Proposed non-convex potential [Hurault et al. '21]:

$$g_\sigma(x) = \frac{1}{2}||x - N_\sigma(x)||^2$$

with a $\mathcal{C}^2$ neural network $N_\sigma : \mathbb{R}^n \to \mathbb{R}^n$ (*e.g.* DRUNet [Zhang et al. '21] with softplux activations).

Training loss :

$$\mathbb{E}_{x,\xi_\sigma}\left[||D_\sigma(x+\xi_\sigma) - x||^2 + \mu \max(||J_{(\mathrm{Id}-D_\sigma)}(x+\xi_\sigma)||_S, 1-\epsilon)\right]$$

✓ D_σ achieves performant denoising.

✓ $\nabla g_\sigma = \mathrm{Id} - D_\sigma$ contractive.

✓ explicit prior

✓ minimization problem

$D_\sigma = \mathrm{Prox}_{\phi_\sigma}$

Prox-PnP

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

Convergence analysis

Objective function :

$$F = f + \phi_\sigma$$

M

x

)

em

tees

Prox-

x^*

✓ explicit prior

✓ minimization problem

$D_\sigma = \text{Prox}_{\phi_\sigma}$

Prox-PnP

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

Convergence analysis [Attouch et al, '13]

Objective function :

$$F = f + \phi_\sigma$$

Prox-PnP-PGD algorithm :

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

For $L_f < 1$, ✗ **Does not allow for low regularization.**

- $F(x_k)$ converges and $||x_{k+1} - x_k|| \to 0$.
- (x_k) converges to a critical point of F.

✓ explicit prior
✓ minimization problem

$D_\sigma = \text{Prox}_{\phi_\sigma}$

Prox-PnP

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

Convergence analysis [Themelis & Patrinos, '20]

Objective function :

$$F = f + \phi_\sigma$$

Prox-PnP-DRS algorithm :

$$x_{k+1} = \frac{1}{2} x_k + \frac{1}{2} (2\,\mathrm{Prox}_{\phi_\sigma} - \mathrm{Id}) \circ (2\,\mathrm{Prox}_f - \mathrm{Id})(x_k)$$

If $Im(f)$ is convex and $L < 1/2$,

- $F(x_k)$ converges and $||x_{k+1} - x_k|| \to 0$.
- (x_k) converges to a critical point of F.

✓ explicit prior
✓ minimization problem

$D_\sigma = \mathrm{Prox}_{\phi_\sigma}$

Prox-PnP

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

Maximum A-Posteriori

$$x^* \in \arg\min_{x \in \mathbb{R}^n} f(x) + g(x)$$

✗ unknown prior

$D_\sigma \approx \text{Prox}_{\tau g}$

Plug-and-Play (PnP)

✓ implicit prior
✓ SOTA restoration
✗ no minimization problem
✗ no convergence guarantees

Prox-Denoiser
$D_\sigma = \text{Prox}_{\phi_\sigma}$

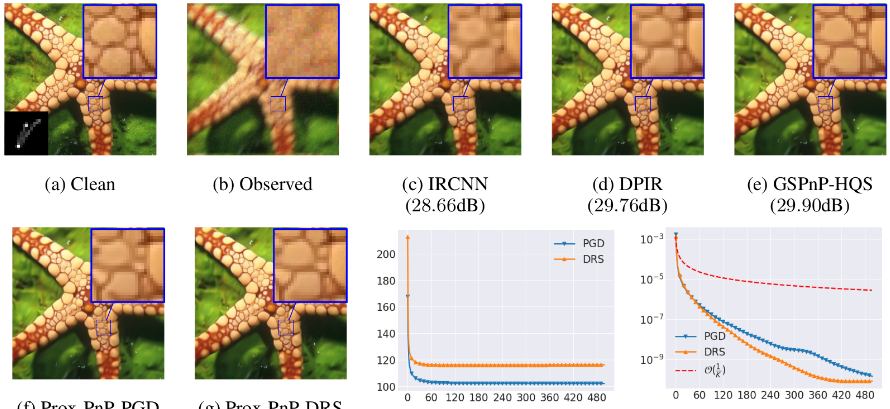
Prox-PnP [Hurault et al, '22]

$$x^* \in \arg\min_{x \in \mathbb{R}^n} f(x) + \phi_\sigma(x)$$

✓ explicit prior
✓ minimization problem
✓ convergence guarantees

Prox-PnP

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

✓ convergence guarantees

Prox-PnP

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

Maximum A-Posteriori

$$x^* \in \arg\min_{x \in \mathbb{R}^n} f(x) + g(x)$$

✗ unknown prior

$D_\sigma \approx \mathrm{Prox}_{\tau g}$

Plug-and-Play (PnP)

✓ implicit prior
✓ SOTA restoration
✗ no minimization problem
✗ no convergence guarantees

Prox-Denoiser
$D_\sigma = \mathrm{Prox}_{\phi_\sigma}$

Prox-PnP [Hurault et al, '22]

$$x^* \in \arg\min_{x \in \mathbb{R}^n} f(x) + \phi_\sigma(x)$$

✓ explicit prior
✓ SOTA restoration
✓ minimization problem
✓ convergence guarantees