Lecture 10: Bridging optimisation & learning: plug & play approaches

Luca Calatroni
CR CNRS, Laboratoire I3S
CNRS, UCA, Inria SAM, France

MSc DSAI - UCA

Inverse problems in image processing

March 24 2023

Image denoisers

Image denoising

Goal: find a noise-free version $x \in \mathbb{R}^n$ of a noisy image $y \in \mathbb{R}^n$ s.t.:

$$y = x + b, \qquad b \sim \mathcal{N}(0, \sigma^2 \boldsymbol{Id})$$

where b is additive, white, Gaussian noise component.

Composite problem:

$$x^* \in \underset{x \in \mathbb{R}^n}{\arg\min} \; \underbrace{\frac{1}{2}\|x - y\|^2}_{f(x)} + \lambda g(x)$$

- Quadratic data term models Gaussian noise. It can be derived by MAP/maximum-likelihood estimation (see first lecture)
- $g(x)$ is the regularisation term enforcing prior information (sparsity, smoothness, gradient smoothness...)
- $\lambda > 0$ controls fidelity VS. regularisation

Image denoising

Goal: find a noise-free version $x \in \mathbb{R}^n$ of a noisy image $y \in \mathbb{R}^n$ s.t.:

$$y = x + b, \qquad b \sim \mathcal{N}(0, \sigma^2 \boldsymbol{Id})$$

where b is additive, white, Gaussian noise component.

Composite problem:

$$x^* \in \operatorname*{arg\,min}_{x \in \mathbb{R}^n} \ \underbrace{\frac{1}{2}\|x - y\|^2}_{f(x)} + \lambda g(x)$$

- Quadratic data term models Gaussian noise. It can be derived by MAP/maximum-likelihood estimation (see first lecture)
- $g(x)$ is the regularisation term enforcing prior information (sparsity, smoothness, gradient smoothness. . .)
- $\lambda > 0$ controls fidelity VS. regularisation

Observe that:

$$x^* \in \operatorname*{arg\,min}_{x \in \mathbb{R}^n} \ \frac{1}{2}\|x - y\|^2 + \lambda g(x) = \operatorname*{arg\,min}_{x \in \mathbb{R}^n} \ \frac{1}{2\lambda}\|x - y\|^2 + g(x) = \operatorname{prox}_{\lambda g}(y)$$

Proximal operators as image denoisers

Goal: given $A \in \mathbb{R}^{m \times n}$, find $x \in \mathbb{R}^n$ from $y \in \mathbb{R}^m$ s.t.:

$$y = Ax + b, \qquad n \sim \mathcal{N}(0, \sigma^2 \boldsymbol{Id})$$

where b is additive, white, Gaussian noise component.

Consider now:

$$x^* \in \arg\min_{x \in \mathbb{R}^n} \ \frac{1}{2}\|Ax - y\|^2 + \lambda g(x)$$

Forward-backward scheme

x^0, $\tau \leq 1/L$, iterate for $k \geq 0$:

$$\begin{aligned} x^{k+1} &= \mathrm{prox}_{\tau\lambda g}(x^k - \tau A^T(Ax^k - y)) \\ &= \arg\min_x \ g(x) + \frac{1}{2\tau\lambda}\|x - \left(x^k - \tau A^T(Ax^k - y)\right)\|^2 = \mathcal{D}(x^k - \tau A^T(Ax^k - y)) \end{aligned}$$

- Interpret proximal steps as a **denoisers** $\mathcal{D}$ (i.e. operators) of gradient iterations
- By definition of the proximal operator (i.e. appearance of $\| \cdot - \cdot \|^2$), assume **noise is Gaussian**

The regularisation function $g(x)$ is called *only* through its prox...

P&P idea

Plug & play methods: idea

Idea: replace $\text{prox}_{\tau\lambda g}$ by some off-the-shelf (black-box) denoiser $\mathcal{D} : \mathbb{R}^n \to \mathbb{R}^n$

Denoiser $\mathcal{D}$: anything doing this job (BM3D, your favourite PhotoShop function...)

Plug & play forward-backward, a.k.a. PnP-ISTA

x^0, $\tau \leq 1/L$, iterate for $k \geq 0$:

$$x^{k+1} = \mathcal{D}(x^k - \tau A^T(Ax^k - y))$$

- No need to define $g(x)$/tune λ! Any denoiser $\mathcal{D}$ (of additive, white Gaussian noise) would work
- $g(x)$ is implicitly defined by the action of the denoiser

Idea: given training set $\{(\tilde{x}_i, y_i)\}_{i=1}^N$ of noise-free/noisy images, train a CNN denoiser $\mathcal{D} : \mathbb{R}^n \to \mathbb{R}^n$ to perform

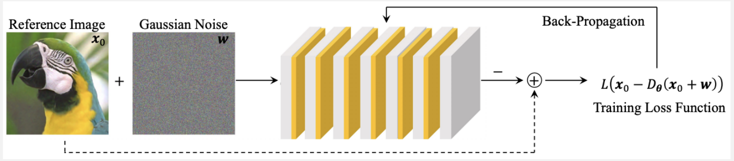
$$\mathcal{D}(y_i) = \tilde{x}_i, \quad \forall i = 1, \ldots, N$$

Integrating physical modelling in DL

Recall that A relates to the physical acquisition model (optical blur, downsampling).

$$x^{k+1} = \mathcal{D}(x^k - \tau A^T(Ax^k - y))$$

PnP-ISTA incorporates naturally **consistency** with the physics/acquired data.



Learning/using a CNN denoiser $\mathcal{D}_\theta$ on super-resolution problems, $A = SH \in \mathbb{R}^{m \times n}$.

Combining with DL

Idea: given training set $\{(\tilde{x}_i, y_i)\}_{i=1}^N$ of noise-free/noisy images, train a CNN denoiser $\mathcal{D} : \mathbb{R}^n \to \mathbb{R}^n$ to perform

$$\mathcal{D}(y_i) = \tilde{x}_i, \quad \forall i = 1, \ldots, N$$

Integrating physical modelling in DL

Recall that A relates to the physical acquisition model (optical blur, downsampling).

$$x^{k+1} = \mathcal{D}(x^k - \tau A^T(Ax^k - y))$$

PnP-ISTA incorporates naturally **consistency** with the physics/acquired data.

Learning/using a CNN denoiser $\mathcal{D}_\theta$ on super-resolution problems, $A = SH \in \mathbb{R}^{m \times n}$.

We thus have an interpretation:

composite optimisation/proximal algorithms $\Rightarrow$ denoisers $\mathcal{D}$

But the question is rather:

given a denoiser $\mathcal{D}$, is there a composite optimisation problem related to $\mathcal{D}$?

$$\exists \quad \arg\min_x \; F(x) := f(x) + g_{\mathcal{D}}(x) \quad ???$$

- Existence of $g_{\mathcal{D}}$? Convergence of $F(x)$?
- Given $\mathcal{D}$, is it the gradient/prox of some function?
- No convexity expected in general

PnP approaches with guarantees

Regularisation by denoising (RED)

Smooth case: given $x_0 \in \mathbb{R}^n$ and $\tau \leq 1/L$, iterate

$$x^{k+1} = x^k - \tau H(x^k), \qquad H(x^k) := \nabla f(x^k) + \lambda(x^k - \mathcal{D}(x^k))$$

where $\nabla f(x^k) = \nabla\left(\frac{1}{2}\|Ax - y\|^2\right) = A^T(Ax^k - y)$

Under restrictive conditions on $\mathcal{D}$ (symmetric Jacobian) $\lambda(x^k - \mathcal{D}(x^k))$ corresponds to the gradient of a function ∇g defined by

$$g(x) = \frac{\lambda}{2} x^T (x - \mathcal{D}(x))$$

Proximal denoisers for convergent plug-and-play

Non-smooth case: for convergence guarantees, the denoiser should have the **gradient-step** structure:

$$\mathcal{D}_\sigma(x) = \nabla\left(\frac{1}{2}\|x\|^2 - f_\sigma(x)\right) = x - \nabla f_\sigma(x) = (\boldsymbol{Id} - \nabla f_\sigma)(x)$$

where $f_\sigma : \mathbb{R}^n \to \mathbb{R}$ is a scalar function parametrised by a NN and $\mathcal{D}_\sigma$ is trained to denoise images with Gaussian noise of level σ.

Theorem (proximal structure)

- If $\frac{1}{2}\|x\|^2 - f_\sigma(x)$ is convex, then there exists $g_\sigma : \mathbb{R}^n \to \mathbb{R} \cup +\{\infty\}$ s.t.:

$$\forall x \in \mathbb{R}^n, \quad \mathcal{D}_\sigma(x) \in \text{prox}_{g_\sigma}(x) \qquad \text{(multi-valued)}$$

- If ∇f_σ is L-Lipschitz with $L < 1$, then $\mathcal{D}_\sigma$ is injective and $\mathcal{D}_\sigma(x) = \text{prox}_{g_\sigma}(x)$, i.e. prox_{g_σ} is **single-valued**.

Note: g_σ is not necessarily convex, but its proximal operator is single-valued.

Convergence of PnP forward-backward

It is thus possible to define a non-convex g_σ such that $\mathrm{prox}_{g_\sigma} = \mathcal{D}_\sigma$. For $\alpha > 0$ regularisation parameter, we now target the function:

$$F(x) := \alpha f(x) + g_\sigma(x)$$

PnP forward-backward

$x^0 \in \mathbb{R}^n$, $\tau = 1$, minimise F by iterating for $k \geq 0$

$$\begin{aligned} z_{k+1} &= x_k - \alpha \nabla f(x_k) \\ x_{k+1} &= \mathcal{D}_\sigma(z_{k+1}) \qquad \text{(PnP-proximal step)} \end{aligned}$$

Theorem (convergence of PnP forward-backward)

If g_σ is C^2 with L-Lipschitz gradient with $L < 1$ and f is smooth with L_f Lipschitz gradient. Then, for $\alpha L_f < 1$, there holds:

- $F(x_k)$ is non-increasing and convergent
- All cluster points of (x_k) are stationary points of F
- Under suitable further assumptions on f and g_σ, the sequence (x_k) converges to a stationary point of F.

Bibliography

. . . very active research area! Contact me if interested!

- S.V. Venkatakrishnan, C.A. Bouman, and B. Wohlberg, *Plug-and-Play Priors for Model Based Reconstruction*, GlobalSIP, 2013.
- Romano, Y., Elad, M., and Milanfar, P., *The little engine that could: Regularization by denoising (RED)*, SIAM Journal on Imaging Sciences, 10(4):1804-1844, 2017.
- Reehorst, E. T. and Schniter, P., *Regularization by denoising: Clarifications and new interpretations*, IEEE transactions on Computational Imaging, 5(1):52-67, 2018.
- Kamilov, U. S., Bouman, C. A., Buzzart, G. T. and Wohlberg, B., *Plug-and-Play Methods for Integrating Physical and Learned Models in Computational Imaging*, IEEE Signal Processing Magazine, 40(1), 2023. Project page
- Hurault, S., Leclaire, A. and Papadakis, N., *Proximal Denoiser for Convergent Plug-and-Play Optimization with Nonconvex Regularization*, International Conference on Machine Learning, 2022.

Questions?

calatroni@i3s.unice.fr