Bregman-based approaches for parameter learning in variational imaging

Kristian Bredies   Enis Chenchene   Alireza Hosseini

Department of Mathematics and Scientific Computing
University of Graz

Signal, Image, Geometry, Modeling, Approximation (SIGMA) 2024

October 30, 2024

Outline

Outline

Variational approaches in imaging

Common setting

$$\min_u \; S_f(u) + R_\alpha(u)$$

- S_f fidelity functional for data f
- R_α regularization functional with parameter α

Variational approaches in imaging

Common setting

$$\min_u \; S_f(u) + R_\alpha(u)$$

- S_f fidelity functional for data f
- R_α regularization functional with parameter α

Examples

- L^2-TV-denoising: $$\min_u \overbrace{\tfrac{1}{2}\|u - f\|_2^2}^{S_f(u)} + \overbrace{\alpha \, \mathrm{TV}(u)}^{R_\alpha(u)}$$

[Rudin/Osher/Fatemi '92]

Variational approaches in imaging

Common setting

$$\min_u \; S_f(u) + R_\alpha(u)$$

- S_f fidelity functional for data f
- R_α regularization functional with parameter α

Examples

- L^2-TV-denoising: $$\min_u \overbrace{\tfrac{1}{2}\|u - f\|_2^2}^{S_f(u)} + \overbrace{\alpha \, \mathrm{TV}(u)}^{R_\alpha(u)}$$

[Rudin/Osher/Fatemi '92]

- $S_f(u) = \frac{1}{2}\|Ku - f\|^2 \rightsquigarrow$ image reconstruction
- $R_\alpha(u) = \int_\Omega \alpha \; \mathrm{d}|\nabla u| \rightsquigarrow$ weighted total variation
- $R_\alpha(u) = \alpha_1 \mathrm{TV}(u) + \alpha_2 \mathrm{TV}^2(u) \rightsquigarrow$ mixed-order TV

Variational approaches in imaging

Common setting

$$\min_u \; S_f(u) + R_\alpha(u)$$

- S_f fidelity functional for data f
- R_α regularization functional with parameter α

Examples

- L^2-TV-denoising: $$\min_u \overbrace{\tfrac{1}{2}\|u-f\|_2^2}^{S_f(u)} + \overbrace{\alpha\,\mathrm{TV}(u)}^{R_\alpha(u)}$$

[Rudin/Osher/Fatemi '92]

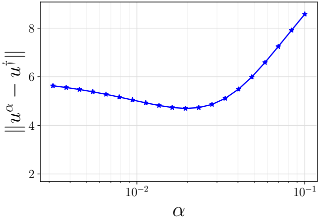
- $S_f(u) = \frac{1}{2}\|Ku - f\|^2 \rightsquigarrow$ image reconstruction
- $R_\alpha(u) = \int_\Omega \alpha \; \mathrm{d}|\nabla u| \rightsquigarrow$ weighted total variation
- $R_\alpha(u) = \alpha_1 \mathrm{TV}(u) + \alpha_2 \mathrm{TV}^2(u) \rightsquigarrow$ mixed-order TV

Parameter choice: How to find α?

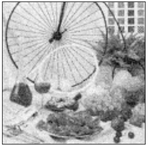
Data-driven parameter choice

Observation

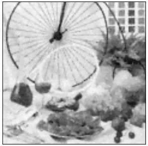
- Different data require different parameters



Ground-truth



Noisy

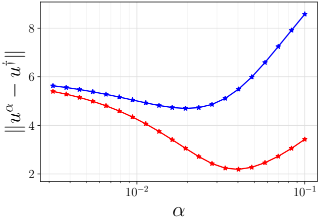


Denoised

Data-driven parameter choice

Observation

- Different data require different parameters



Ground-truth

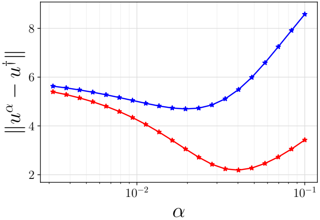
Noisy

Denoised

Data-driven parameter choice

Observation

- Different data require different parameters




Ground-truth


Noisy


Denoised

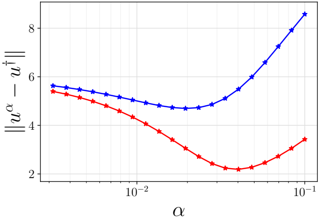
Model

- $\alpha : Y \to A$ parameter choice function
- Y data space, A parameter space
- Solve $$\min_u \; S_f(u) + R_{\alpha(f)}(u)$$

Data-driven parameter choice

Observation

- Different data require different parameters



Ground-truth

Noisy

Denoised

Model

- $\alpha : Y \to A$ parameter choice function
- Y data space, A parameter space
- Solve $$\min_u \; S_f(u) + R_{\alpha(f)}(u)$$

$\rightsquigarrow$ **Obtain α using learning strategies**

Bilevel parameter learning

First idea Given

1. a **training set** $\{(u_i^\dagger, f_i)\}_{i=1}^{N_t}$ of *ground-truth/noisy* images,
2. a **prediction model** $\mathcal{F} \subset \{\alpha : Y \to A\}$,

solve the bilevel problem

$$\begin{cases} \min\limits_{\alpha \in \mathcal{F}} \dfrac{1}{2N_t} \displaystyle\sum_{i=1}^{N_t} \|u_i^\dagger - u_i^{\alpha(f_i)}\|^2, \\ u_i^{\alpha(f_i)} \in \arg\min\limits_u S_{f_i}(u) + R_{\alpha(f_i)}(u) \quad \text{for all } i \end{cases}$$

[De los Reyes/Schönlieb/Valkonen '17], [Kunisch/Pock '12]

Bilevel parameter learning

First idea Given

1. a **training set** $\{(u_i^\dagger, f_i)\}_{i=1}^{N_t}$ of *ground-truth/noisy* images,
2. a **prediction model** $\mathcal{F} \subset \{\alpha : Y \to A\}$,

solve the bilevel problem

$$\begin{cases} \min\limits_{\alpha \in \mathcal{F}} \ \dfrac{1}{2N_t} \displaystyle\sum_{i=1}^{N_t} \|u_i^\dagger - u_i^{\alpha(f_i)}\|^2, \\ u_i^{\alpha(f_i)} \in \arg\min\limits_u S_{f_i}(u) + R_{\alpha(f_i)}(u) \quad \text{for all } i \end{cases}$$

[De los Reyes/Schönlieb/Valkonen '17], [Kunisch/Pock '12]

$\rightsquigarrow$ **difficult to solve numerically**

Bilevel parameter learning

First idea Given

1. a **training set** $\{(u_i^\dagger, f_i)\}_{i=1}^{N_t}$ of *ground-truth/noisy* images,
2. a **prediction model** $\mathcal{F} \subset \{\alpha : Y \to A\}$,

solve the bilevel problem

$$\begin{cases} \min\limits_{\alpha \in \mathcal{F}} \dfrac{1}{2N_t} \sum\limits_{i=1}^{N_t} \|u_i^\dagger - u_i^{\alpha(f_i)}\|^2, \\ u_i^{\alpha(f_i)} \in \arg\min\limits_{u} S_{f_i}(u) + R_{\alpha(f_i)}(u) \quad \text{for all } i \end{cases}$$

[De los Reyes/Schönlieb/Valkonen '17], [Kunisch/Pock '12]

$\rightsquigarrow$ **difficult to solve numerically**

$\rightsquigarrow$ **employ Bregman relaxation**

Bregman divergences

Definitions

- For $F : X \to \mathbb{R}_\infty$ proper, convex, lsc., $\xi \in X^*$ is a subgradient at u, i.e., $\xi \in \partial F(u)$, if
$$F(u) + \langle \xi, u' - u \rangle \leq F(u') \qquad \text{for all } u' \in X$$
- For $\xi \in \partial F(u)$, the Bregman divergence is
$$\mathcal{D}_F^\xi(u', u) = F(u') - F(u) - \langle \xi, u' - u \rangle$$

Bregman divergences

Definitions

- For $F : X \to \mathbb{R}_\infty$ proper, convex, lsc., $\xi \in X^*$ is a subgradient at u, i.e., $\xi \in \partial F(u)$, if
$$F(u) + \langle \xi, u' - u \rangle \leq F(u') \qquad \text{for all } u' \in X$$
- For $\xi \in \partial F(u)$, the Bregman divergence is
$$\mathcal{D}_F^\xi(u', u) = F(u') - F(u) - \langle \xi, u' - u \rangle$$

Examples

- $F(u) = \frac{1}{2}\|u - f\|_2^2 \rightsquigarrow \mathcal{D}_F^{u-f}(u', u) = \frac{1}{2}\|u' - u\|_2^2$
- $F(u) = \frac{1}{2}\|Ku - f\|_2^2 \rightsquigarrow \mathcal{D}_F^{K^*(Ku-f)}(u', u) = \frac{1}{2}\|Ku' - Ku\|_2^2$

Fenchel–Rockafellar duality

Dual functional

- For $F : X \to \mathbb{R}_\infty$ proper, convex, lsc., the Fenchel dual F^* is

$$F^*(u^*) = \sup_u \langle u^*, u \rangle - F(u) \qquad \text{for all } u^* \in X^*$$

Fenchel–Rockafellar duality

Dual functional

- For $F : X \to \mathbb{R}_\infty$ proper, convex, lsc., the Fenchel dual F^* is

$$F^*(u^*) = \sup_u \langle u^*, u\rangle - F(u) \qquad \text{for all } u^* \in X^*$$

Fenchel–Rockafellar duality

- For $F : X \to \mathbb{R}_\infty$, $G : Y \to \mathbb{R}_\infty$ proper, convex, lsc., $K : X \to Y$ linear and bounded, we have

$$\min_u \; F(u) + G(Ku) = \max_v \; -F^*(-K^*v) - G^*(v)$$

under qualification conditions

Fenchel–Rockafellar duality

Dual functional

- For $F : X \to \mathbb{R}_\infty$ proper, convex, lsc., the Fenchel dual F^* is
$$F^*(u^*) = \sup_u \langle u^*, u\rangle - F(u) \qquad \text{for all } u^* \in X^*$$

Fenchel–Rockafellar duality

- For $F : X \to \mathbb{R}_\infty$, $G : Y \to \mathbb{R}_\infty$ proper, convex, lsc., $K : X \to Y$ linear and bounded, we have
$$\min_u \; F(u) + G(Ku) = \max_v \; -F^*(-K^*v) - G^*(v)$$
under qualification conditions

Primal-dual gap

$$\mathcal{G}(u, v) = F(u) + G(Ku) + F^*(-K^*v) + G^*(v)$$

Bregman primal-dual gap

Primal-dual optimality system

$$-K^*v^* \in \partial F(u^*), \qquad Ku^* \in \partial G^*(v^*)$$

if and only if (u^*, v^*) is a primal-dual solution pair

Bregman primal-dual gap

Primal-dual optimality system

$$-K^* v^* \in \partial F(u^*), \qquad Ku^* \in \partial G^*(v^*)$$

if and only if (u^*, v^*) is a primal-dual solution pair

Proposition *[B./Chenchene/Hosseini '23]*

$$\mathcal{G}(u, v) = \mathcal{D}_F^{-K^*v^*}(u, u^*) + \mathcal{D}_{F^*}^{u^*}(-K^*v, -K^*v^*) + \mathcal{D}_G^{v^*}(Ku, Ku^*) + \mathcal{D}_{G^*}^{Ku^*}(v, v^*)$$

Bregman primal-dual gap

Proof

$$\mathcal{G}(u, v) = F(u) + G(Ku) + F^*(-K^*v) + G^*(v)$$

Bregman primal-dual gap

Proof

$$\begin{aligned}\mathcal{G}(u,v) = F(u) + G(Ku) + F^*(-K^*v) + G^*(v) \\ - F(u^*) - G(Ku^*) - F^*(-K^*v^*) - G^*(v^*)\end{aligned}$$

Bregman primal-dual gap

Proof

$$
\begin{aligned}
\mathcal{G}(u,v) = {} & F(u) + G(Ku) + F^*(-K^*v) + G^*(v) \\
& - F(u^*) - G(Ku^*) - F^*(-K^*v^*) - G^*(v^*) \\
& - \langle -K^*v^*, u - u^* \rangle - \langle u^*, -K^*v + K^*v^* \rangle \\
& - \langle v^*, Ku - Ku^* \rangle - \langle Ku^*, v - v^* \rangle
\end{aligned}
$$

Bregman primal-dual gap

Proof

$$
\begin{aligned}
\mathcal{G}(u,v) &= F(u) + G(Ku) + F^*(-K^*v) + G^*(v) \\
&\quad - F(u^*) - G(Ku^*) - F^*(-K^*v^*) - G^*(v^*) \\
&\quad - \langle -K^*v^*, u - u^* \rangle - \langle u^*, -K^*v + K^*v^* \rangle \\
&\quad - \langle v^*, Ku - Ku^* \rangle - \langle Ku^*, v - v^* \rangle \\
&= \mathcal{D}_F^{-K^*v^*}(u,u^*) + \mathcal{D}_{F^*}^{u^*}(-K^*v, -K^*v^*) \\
&\quad + \mathcal{D}_G^{v^*}(Ku, Ku^*) + \mathcal{D}_{G^*}^{Ku^*}(v, v^*)
\end{aligned}
$$

Bregman primal-dual gap

Proof

$$
\begin{aligned}
\mathcal{G}(u,v) &= F(u) + G(Ku) + F^*(-K^*v) + G^*(v) \\
&\quad - F(u^*) - G(Ku^*) - F^*(-K^*v^*) - G^*(v^*) \\
&\quad - \langle -K^*v^*, u - u^* \rangle - \langle u^*, -K^*v + K^*v^* \rangle \\
&\quad - \langle v^*, Ku - Ku^* \rangle - \langle Ku^*, v - v^* \rangle \\
&= \mathcal{D}_F^{-K^*v^*}(u,u^*) + \mathcal{D}_{F^*}^{u^*}(-K^*v, -K^*v^*) \\
&\quad + \mathcal{D}_G^{v^*}(Ku, Ku^*) + \mathcal{D}_{G^*}^{Ku^*}(v, v^*)
\end{aligned}
$$

□

Bregman primal-dual gap

Proof

$$\begin{aligned}\mathcal{G}(u,v) &= F(u) + G(Ku) + F^*(-K^*v) + G^*(v) \\ &\quad - F(u^*) - G(Ku^*) - F^*(-K^*v^*) - G^*(v^*) \\ &\quad - \langle -K^*v^*, u - u^*\rangle - \langle u^*, -K^*v + K^*v^*\rangle \\ &\quad - \langle v^*, Ku - Ku^*\rangle - \langle Ku^*, v - v^*\rangle \\ &= \mathcal{D}_F^{-K^*v^*}(u,u^*) + \mathcal{D}_{F^*}^{u^*}(-K^*v, -K^*v^*) \\ &\quad + \mathcal{D}_G^{v^*}(Ku, Ku^*) + \mathcal{D}_{G^*}^{Ku^*}(v, v^*)\end{aligned}$$

□

Consequence

$$\min_v \ \mathcal{G}(u,v) = \mathcal{D}_F^{-K^*v^*}(u,u^*) + \mathcal{D}_G^{v^*}(Ku, Ku^*)$$

Bregman primal-dual gap

Proof

$$\begin{aligned}\mathcal{G}(u,v) &= F(u) + G(Ku) + F^*(-K^*v) + G^*(v)\\ &\quad - F(u^*) - G(Ku^*) - F^*(-K^*v^*) - G^*(v^*)\\ &\quad - \langle -K^*v^*, u - u^*\rangle - \langle u^*, -K^*v + K^*v^*\rangle\\ &\quad - \langle v^*, Ku - Ku^*\rangle - \langle Ku^*, v - v^*\rangle\\ &= \mathcal{D}_F^{-K^*v^*}(u,u^*) + \mathcal{D}_{F^*}^{u^*}(-K^*v, -K^*v^*)\\ &\quad + \mathcal{D}_G^{v^*}(Ku,Ku^*) + \mathcal{D}_{G^*}^{Ku^*}(v,v^*)\end{aligned}$$

$\square$

Consequence

$$\min_v \mathcal{G}(u,v) = \mathcal{D}_F^{-K^*v^*}(u,u^*) + \mathcal{D}_G^{v^*}(Ku,Ku^*)$$

$\rightsquigarrow$ **use as loss function**

Bregman loss for parameter learning

- With $F = S_f$, $G = R_\alpha$ and K identity
- $\mathcal{G}_\alpha$ duality gap associated with $\min\ S_f + R_\alpha$

Bregman loss

$$\min_v \ \mathcal{G}_\alpha(u, v) = \mathcal{D}_{S_f}^{-v^*}(u, u^*) + \mathcal{D}_{R_\alpha}^{v^*}(u, u^*)$$

- $\mathcal{D}_{S_f}^{-v^*}(u, u^*) \sim$ data fidelity loss, e.g.,

$$S_f(u) = \tfrac{1}{2}\|u - f\|_2^2 \rightsquigarrow \mathcal{D}_{S_f}^{-v^*}(u, u^*) = \tfrac{1}{2}\|u - u^*\|_2^2$$

- $\mathcal{D}_{R_\alpha}^{v^*}(u, u^*) \sim$ Bregman relaxation w.r.t. regularizer

Bregman-relaxed parameter learning

Original bilevel problem

$$
\begin{cases}
\min_{\alpha \in \mathcal{F}} \dfrac{1}{N_t} \displaystyle\sum_{i=1}^{N_t} \tfrac{1}{2} \| u_i^\dagger - u_i^{\alpha(f_i)} \|^2 & , \\
u_i^{\alpha(f_i)} \in \arg\min_u S_{f_i}(u) + R_{\alpha(f_i)}(u) \quad \text{for all } i &
\end{cases}
$$

Bregman-relaxed parameter learning

Bregman-relaxed problem

$$\begin{cases} \min\limits_{\alpha \in \mathcal{F}} \dfrac{1}{N_t} \displaystyle\sum_{i=1}^{N_t} \tfrac{1}{2} \| u_i^\dagger - u_i^{\alpha(f_i)} \|^2 + \mathcal{D}_{R_{\alpha(f_i)}}^{v_i^{\alpha(f_i)}}(u_i^\dagger, u_i^{\alpha(f_i)}), \\ u_i^{\alpha(f_i)} \in \arg\min\limits_u S_{f_i}(u) + R_{\alpha(f_i)}(u) \quad \text{for all } i, \\ v_i^{\alpha(f_i)} \in \arg\min\limits_v S_{f_i}^*(-v) + R_{\alpha(f_i)}^*(v) \quad \text{for all } i \end{cases}$$

in the denoising case $S_{f_i}(u) = \frac{1}{2}\|u - f_i\|_2^2$

Bregman-relaxed parameter learning

Bregman-relaxed problem

$$
\begin{cases}
\min\limits_{\alpha \in \mathcal{F}} \dfrac{1}{N_t} \displaystyle\sum_{i=1}^{N_t} \tfrac{1}{2} \| u_i^\dagger - u_i^{\alpha(f_i)} \|^2 + \mathcal{D}_{R_{\alpha(f_i)}}^{v_i^{\alpha(f_i)}}(u_i^\dagger, u_i^{\alpha(f_i)}), \\
u_i^{\alpha(f_i)} \in \arg\min\limits_u S_{f_i}(u) + R_{\alpha(f_i)}(u) \quad \text{for all } i, \\
v_i^{\alpha(f_i)} \in \arg\min\limits_v S_{f_i}^*(-v) + R_{\alpha(f_i)}^*(v) \quad \text{for all } i
\end{cases}
$$

in the denoising case $S_{f_i}(u) = \frac{1}{2}\|u - f_i\|_2^2$

Equivalent monolevel problem

$$
\min_{\alpha \in \mathcal{F}, v_1, \ldots, v_{N_t}} \frac{1}{N_t} \sum_{i=1}^{N_t} \mathcal{G}_{\alpha(f_i)}(u_i^\dagger, v_i)
$$

Convexity of relaxed problem

Bregman-relaxed problem

$$\min_{\alpha\in\mathcal{F},v_1,\ldots,v_{N_t}} \frac{1}{N_t}\sum_{i=1}^{N_t} \mathcal{G}_{\alpha(f_i)}(u_i^\dagger, v_i)$$

- Objective is convex in many situations

Convexity of relaxed problem

Bregman-relaxed problem

$$\min_{\alpha\in\mathcal{F},v_1,\ldots,v_{N_t}} \frac{1}{N_t}\sum_{i=1}^{N_t} \mathcal{G}_{\alpha(f_i)}(u_i^\dagger, v_i)$$

- Objective is convex in many situations

Example

- $\mathcal{F} = \{(\alpha_1\mathbf{1},\ldots,\alpha_M\mathbf{1}) \mid \alpha_1,\ldots,\alpha_M \geq 0\}$
- $R_\alpha(u) = \sum_{j=1}^M \alpha_j|u|_j$, $|u|_j \sim$ semi-norm regularizer

Convexity of relaxed problem

Bregman-relaxed problem

$$\min_{\alpha \in \mathcal{F}, v_1, \ldots, v_{N_t}} \frac{1}{N_t} \sum_{i=1}^{N_t} \mathcal{G}_{\alpha(f_i)}(u_i^\dagger, v_i)$$

- Objective is convex in many situations

Example

- $\mathcal{F} = \{(\alpha_1 \mathbf{1}, \ldots, \alpha_M \mathbf{1}) \mid \alpha_1, \ldots, \alpha_M \geq 0\}$
- $R_\alpha(u) = \sum_{j=1}^{M} \alpha_j |u|_j$, $|u|_j \sim$ semi-norm regularizer

$$\mathcal{G}_{\alpha(f_i)}(u_i^\dagger, v_i) = \min_{\substack{w_{1,i}+\ldots+w_{M,i}=v_i, \\ |w_{j,i}|_{j*} \leq \alpha_j}} \tfrac{1}{2}\|u_i^\dagger - f_i\|_2^2 + \sum_{j=1}^{M} \alpha_j |u_i^\dagger|_j + \tfrac{1}{2}\|f_i + v_i\|_2^2 - \tfrac{1}{2}\|f_i\|_2^2$$

- $|w_{j,i}|_{j*} \sim$ dual semi-norm $\rightsquigarrow$ jointly convex in α and v_i

Bregman learning for TV denoising

Discrete TV denoising

Solve

$$u^\alpha = \operatorname*{arg\,min}_{u \in \mathbb{R}^n} \ \tfrac{1}{2}\|u - f\|^2 + \alpha \, \mathrm{TV}(u),$$

$\mathrm{TV}(u) = \|\nabla u\|_{1,2}$ is a discrete total variation, $\alpha \geq 0$

Bregman learning for TV denoising

Discrete TV denoising

Solve

$$u^\alpha = \arg\min_{u \in \mathbb{R}^n} \ \tfrac{1}{2}\|u - f\|^2 + \alpha \, \mathrm{TV}(u),$$

$\mathrm{TV}(u) = \|\nabla u\|_{1,2}$ is a discrete total variation, $\alpha \geq 0$

Primal-dual gap

$$\begin{aligned} \mathcal{G}_\alpha(u, v) = \tfrac{1}{2}\|u - f\|_2^2 + \alpha \, \mathrm{TV}(u) + \tfrac{1}{2}\|\operatorname{div} v + f\|_2^2 \\ - \tfrac{1}{2}\|f\|_2^2 + I_{\{\|\cdot\|_{\infty,2} \leq \alpha\}}(v) \end{aligned}$$

Bregman learning for TV denoising

Discrete TV denoising

Solve

$$u^\alpha = \arg\min_{u \in \mathbb{R}^n} \ \tfrac{1}{2}\|u - f\|^2 + \alpha \operatorname{TV}(u),$$

$\operatorname{TV}(u) = \|\nabla u\|_{1,2}$ is a discrete total variation, $\alpha \geq 0$

Primal-dual gap

$$\mathcal{G}_\alpha(u, v) = \tfrac{1}{2}\|u - f\|_2^2 + \alpha \operatorname{TV}(u) + \tfrac{1}{2}\|\operatorname{div} v + f\|_2^2 - \tfrac{1}{2}\|f\|_2^2 + I_{\{\|\cdot\|_{\infty,2} \leq \alpha\}}(v)$$

Equivalent learning problem

$$\min_{\substack{\alpha \in \mathcal{F}, \\ \|v_i\|_{\infty,2} \leq \alpha(f_i)}} \ \frac{1}{N_t} \sum_{i=1}^{N_t} \tfrac{1}{2}\|\operatorname{div} v_i + f_i\|_2^2 + \alpha(f_i) \operatorname{TV}(u_i^\dagger)$$

A quadratic prediction model

$$\min_{\substack{\alpha \in \mathcal{F} \\ \forall i:\ \|v_i\|_{\infty,2} \leq \alpha(f_i)}} \frac{1}{N_t} \sum_{i=1}^{N_t} \left(\frac{1}{2} \| \operatorname{div} v_i + f_i \|^2 + \alpha(f_i) \operatorname{TV}(u_i^\dagger) \right)$$

A quadratic prediction model

$$\min_{\substack{\alpha\in\mathcal{F} \\ \forall i:\ \|v_i\|_{\infty,2}\leq\alpha(f_i)}} \frac{1}{N_t}\sum_{i=1}^{N_t}\left(\frac{1}{2}\|\operatorname{div} v_i + f_i\|^2 + \alpha(f_i)\,\mathrm{TV}(u_i^\dagger)\right)$$

Quadratic prediction model

- $\mathcal{F} = \{f \mapsto \bar{f}^* A \bar{f} \mid A \succcurlyeq 0\}$ quadratic polynomials ≥ 0
- $\bar{f} = (f, 1)$ data with bias

A quadratic prediction model

$$\min_{\substack{\alpha \in \mathcal{F} \\ \forall i:\ \|v_i\|_{\infty,2} \leq \alpha(f_i)}} \frac{1}{N_t} \sum_{i=1}^{N_t} \left(\frac{1}{2} \| \operatorname{div} v_i + f_i \|^2 + \alpha(f_i) \operatorname{TV}(u_i^\dagger) \right)$$

Quadratic prediction model

- $\mathcal{F} = \{ f \mapsto \bar{f}^* A \bar{f} \mid A \succcurlyeq 0 \}$ quadratic polynomials ≥ 0
- $\bar{f} = (f, 1)$ data with bias

Convex TV parameter learning

$$\min_{(A, \boldsymbol{v}) \in \mathcal{C}} \frac{1}{N_t} \sum_{i=1}^{N_t} \left(\frac{1}{2} \| \operatorname{div} v_i + f_i \|^2 + \bar{f}_i^* A \bar{f}_i \operatorname{TV}(u_i^\dagger) \right),$$

$$\mathcal{C} = \{ (A, \boldsymbol{v} = (v_1, \ldots v_{N_t})) \,|\, A \succcurlyeq 0, \ \|v_i\|_{\infty,2} \leq \bar{f}_i^* A \bar{f}_i \text{ for all } i \}$$

A quadratic prediction model

$$\min_{\substack{\alpha\in\mathcal{F} \\ \forall i:\ \|v_i\|_{\infty,2}\le\alpha(f_i)}} \frac{1}{N_t}\sum_{i=1}^{N_t}\left(\frac{1}{2}\|\operatorname{div} v_i + f_i\|^2 + \alpha(f_i)\,\mathrm{TV}(u_i^\dagger)\right)$$

Quadratic prediction model

- $\mathcal{F} = \{f \mapsto \bar{f}^* A \bar{f} \mid A \succcurlyeq 0\}$ quadratic polynomials ≥ 0
- $\bar{f} = (f, 1)$ data with bias

Convex TV parameter learning

$$\min_{(A,v)\in\mathcal{C}} \frac{1}{N_t}\sum_{i=1}^{N_t}\left(\frac{1}{2}\|\operatorname{div} v_i + f_i\|^2 + \bar{f}_i^* A \bar{f}_i\,\mathrm{TV}(u_i^\dagger)\right),$$

$$\mathcal{C} = \{(A, v = (v_1, \ldots v_{N_t})) \,|\, A \succcurlyeq 0,\ \|v_i\|_{\infty,2} \le \bar{f}_i^* A \bar{f}_i \text{ for all } i\}$$

Note $\mathcal{C}$ is unbounded and $P_{\mathcal{C}}$ is challenging

Outline

Motivation

Problem $\min_{u \in H} f(u) + g(u)$ $f, g \in \Gamma_0(H)$, f smooth

1. **(Proximal gradient)** Start from u^0 and **iterate**

$$u^{k+1} = u^k + \theta_k(\tilde{u}^k - u^k), \quad \tilde{u}^k = \mathrm{prox}_{\tau_k g}(u^k - \tau_k \nabla f(u^k))$$

[Goldstein '64], [Bruck '77], ...

Motivation

Problem $\min_{u \in H} f(u) + g(u) \qquad f, g \in \Gamma_0(H), f$ smooth

1 **(Proximal gradient)** Start from u^0 and **iterate**

$$u^{k+1} = u^k + \theta_k(\widetilde{u}^k - u^k), \qquad \widetilde{u}^k = \mathrm{prox}_{\tau_k g}(u^k - \tau_k \nabla f(u^k))$$

- Robust convergence guarantees for all $g \in \Gamma_0(H)$

[Goldstein '64], [Bruck '77], ...

Motivation

Problem $\min_{u \in H} f(u) + g(u)$ $\quad f, g \in \Gamma_0(H)$, f smooth

1. **(Proximal gradient)** Start from u^0 and **iterate**

$$u^{k+1} = u^k + \theta_k(\widetilde{u}^k - u^k), \quad \widetilde{u}^k = \mathrm{prox}_{\tau_k g}(u^k - \tau_k \nabla f(u^k))$$

- Robust convergence guarantees for all $g \in \Gamma_0(H)$
- $\mathrm{prox}_{\tau_k g}$ can be costly

[Goldstein '64], [Bruck '77], ...

Motivation

Problem $\min_{u \in H} f(u) + g(u)$ $\quad f, g \in \Gamma_0(H)$, f smooth

2 **(Gen. conditional gradient)** Start from u^0 and **iterate**

$$u^{k+1} = u^k + \theta_k(\tilde{u}^k - u^k), \quad \tilde{u}^k \in \arg\min_{v \in H} \langle \nabla f(u^k), v \rangle + g(v)$$

[Frank/Wolfe, '56], [B./Lorenz '08], ...

Motivation

Problem $\min_{u \in H} \; f(u) + g(u) \qquad f, g \in \Gamma_0(H), f$ smooth

2 **(Gen. conditional gradient)** Start from u^0 and **iterate**

$$u^{k+1} = u^k + \theta_k(\tilde{u}^k - u^k), \quad \tilde{u}^k \in \arg\min_{v \in H} \langle \nabla f(u^k), v \rangle + g(v)$$

- Requires g (strongly) coercive

[Frank/Wolfe, '56], [B./Lorenz '08], ...

Motivation

Problem $\min_{u \in H} f(u) + g(u) \qquad f, g \in \Gamma_0(H), f$ smooth

2 **(Gen. conditional gradient)** Start from u^0 and **iterate**

$$u^{k+1} = u^k + \theta_k(\widetilde{u}^k - u^k), \quad \widetilde{u}^k \in \arg\min_{v \in H} \langle \nabla f(u^k), v\rangle + g(v)$$

- Requires g (strongly) coercive
- Avoids $\mathrm{prox}_{\tau_k g}$

[Frank/Wolfe, '56], [B./Lorenz '08], ...

Motivation

Problem $\min_{u \in H} f(u) + g(u)$ $\quad f, g \in \Gamma_0(H)$, f smooth

2 **(Gen. conditional gradient)** Start from u^0 and **iterate**

$$u^{k+1} = u^k + \theta_k(\tilde{u}^k - u^k), \quad \tilde{u}^k \in \arg\min_{v \in H} \langle \nabla f(u^k), v \rangle + g(v)$$

- Requires g (strongly) coercive
- Avoids $\mathrm{prox}_{\tau_k g}$

[Frank/Wolfe, '56], [B./Lorenz '08], ...

Can we combine the advantages of these two methods in a new hybrid variant? $\rightsquigarrow$ HPGCG

HPGCG: Idea

Problem $\min_{u \in H} f(u) + g(u)$ $\quad f, g \in \Gamma_0(H)$, f smooth

Consider instead *[B./Lorenz/Maass '04]*

$$\min_{u \in H} f(u) - \tfrac{1}{2}\|u\|_P^2 + \tfrac{1}{2}\|u\|_P^2 + g(u),$$

where $\|u\|_P^2 := \langle Pu, u \rangle$ is the semi-norm[2] induced by $P \succcurlyeq 0$

HPGCG: Idea

Problem $\min_{u \in H} f(u) + g(u)$ $\quad f, g \in \Gamma_0(H)$, f smooth

Consider instead *[B./Lorenz/Maass '04]*

$$\min_{u \in H} \underbrace{f(u) - \tfrac{1}{2}\|u\|_P^2}_{\rightsquigarrow \textit{ linearize}} + \tfrac{1}{2}\|u\|_P^2 + g(u),$$

where $\|u\|_P^2 := \langle Pu, u \rangle$ is the semi-norm[2] induced by $P \succcurlyeq 0$

HPGCG: Idea

Problem $\min_{u \in H} \; f(u) + g(u) \qquad f, g \in \Gamma_0(H), f$ smooth

Consider instead *[B./Lorenz/Maass '04]*

$$\min_{u \in H} \; \underbrace{f(u) - \tfrac{1}{2}\|u\|_P^2}_{\rightsquigarrow \textit{ linearize}} + \tfrac{1}{2}\|u\|_P^2 + g(u),$$

where $\|u\|_P^2 := \langle Pu, u\rangle$ is the semi-norm[2] induced by $P \succcurlyeq 0$

Initialize: $u^0 \in \mathrm{dom}(g)$

for $k = 0, 1, \ldots$ **do**

$$\tilde{u}^k \in \arg\min_{v \in H} \; \langle \nabla f(u^k) - Pu^k, v\rangle + \tfrac{1}{2}\|v\|_P^2 + g(v)$$

$$u^{k+1} = u^k + \theta_k(\tilde{u}^k - u^k)$$

end

HPGCG: A hybrid method

If $P \succ 0$, then $u^{k+1} = u^k + \theta_k(\widetilde{u}^k - u^k)$, where:

$$\widetilde{u}^k = \arg\min_{v \in H} \langle \nabla f(u^k) - Pu^k, v \rangle + \frac{1}{2}\|v\|_P^2 + g(v)$$

HPGCG: A hybrid method

If $P \succ 0$, then $u^{k+1} = u^k + \theta_k(\tilde{u}^k - u^k)$, where:

$$\begin{aligned} \tilde{u}^k &= \arg\min_{v \in H} \langle \nabla f(u^k) - Pu^k, v \rangle + \tfrac{1}{2}\|v\|_P^2 + g(v) \\ &= \arg\min_{v \in H} \langle P^{-1}\nabla f(u^k) - u^k, v \rangle_P + \tfrac{1}{2}\|v\|_P^2 + g(v) \end{aligned}$$

HPGCG: A hybrid method

If $P \succ 0$, then $u^{k+1} = u^k + \theta_k(\widetilde{u}^k - u^k)$, where:

$$\begin{aligned}
\widetilde{u}^k &= \arg\min_{v \in H} \ \langle \nabla f(u^k) - Pu^k, v\rangle + \tfrac{1}{2}\|v\|_P^2 + g(v) \\
&= \arg\min_{v \in H} \ \langle P^{-1}\nabla f(u^k) - u^k, v\rangle_P + \tfrac{1}{2}\|v\|_P^2 + g(v) \\
&= \arg\min_{v \in H} \ \tfrac{1}{2}\|v - (u^k - P^{-1}\nabla f(u^k))\|_P^2 + g(v).
\end{aligned}$$

HPGCG: A hybrid method

If $P \succ 0$, then $u^{k+1} = u^k + \theta_k(\widetilde{u}^k - u^k)$, where:

$$\begin{aligned}\widetilde{u}^k &= \arg\min_{v \in H} \langle \nabla f(u^k) - Pu^k, v \rangle + \tfrac{1}{2}\|v\|_P^2 + g(v) \\ &= \arg\min_{v \in H} \langle P^{-1}\nabla f(u^k) - u^k, v \rangle_P + \tfrac{1}{2}\|v\|_P^2 + g(v) \\ &= \arg\min_{v \in H} \tfrac{1}{2}\|v - (u^k - P^{-1}\nabla f(u^k))\|_P^2 + g(v).\end{aligned}$$

- If $P \succ 0$, we have preconditioned **proximal gradient**

HPGCG: A hybrid method

If $P \succ 0$, then $u^{k+1} = u^k + \theta_k(\widetilde{u}^k - u^k)$, where:

$$\begin{aligned}
\widetilde{u}^k &= \arg\min_{v \in H} \langle \nabla f(u^k) - Pu^k, v \rangle + \tfrac{1}{2}\|v\|_P^2 + g(v) \\
&= \arg\min_{v \in H} \langle P^{-1}\nabla f(u^k) - u^k, v \rangle_P + \tfrac{1}{2}\|v\|_P^2 + g(v) \\
&= \arg\min_{v \in H} \tfrac{1}{2}\|v - (u^k - P^{-1}\nabla f(u^k))\|_P^2 + g(v).
\end{aligned}$$

- If $P \succ 0$, we have preconditioned **proximal gradient**
- If $P = 0$, we retrieve **Gen. conditional gradient**

HPGCG: A hybrid method

If $P \succ 0$, then $u^{k+1} = u^k + \theta_k(\tilde{u}^k - u^k)$, where:

$$\begin{aligned} \tilde{u}^k &= \arg\min_{v \in H} \langle \nabla f(u^k) - Pu^k, v \rangle + \tfrac{1}{2}\|v\|_P^2 + g(v) \\ &= \arg\min_{v \in H} \langle P^{-1}\nabla f(u^k) - u^k, v \rangle_P + \tfrac{1}{2}\|v\|_P^2 + g(v) \\ &= \arg\min_{v \in H} \tfrac{1}{2}\|v - (u^k - P^{-1}\nabla f(u^k))\|_P^2 + g(v). \end{aligned}$$

- If $P \succ 0$, we have preconditioned **proximal gradient**
- If $P = 0$, we retrieve **Gen. conditional gradient**
- If $P \succcurlyeq 0 \rightsquigarrow$ hybrid proximal generalized conditional gradient method **(HPGCG)**

HPGCG: Step-size

At $k \in \mathbb{N}$, **HPGCG** requires minimizing over H

$$v \mapsto H_{u^k}(v) := \langle \nabla f(u^k) - Pu^k, v \rangle + \frac{1}{2}\|v\|_P^2 + g(v).$$

Some **remarks**

HPGCG: Step-size

At $k \in \mathbb{N}$, **HPGCG** requires minimizing over H

$$v \mapsto H_{u^k}(v) := \langle \nabla f(u^k) - Pu^k, v \rangle + \tfrac{1}{2}\|v\|_P^2 + g(v).$$

Some **remarks**

- It requires $\frac{1}{2}\|\cdot\|_P^2 + g$ to be (strongly) **coercive**,
  if $g = \mathbb{I}_{\mathcal{C}} \rightsquigarrow$ assumption $\mathcal{C}$ bounded can be dropped

HPGCG: Step-size

At $k \in \mathbb{N}$, **HPGCG** requires minimizing over H

$$v \mapsto H_{u^k}(v) := \langle \nabla f(u^k) - Pu^k, v\rangle + \tfrac{1}{2}\|v\|_P^2 + g(v).$$

Some **remarks**

- It requires $\frac{1}{2}\|\cdot\|_P^2 + g$ to be (strongly) **coercive**,
  if $g = \mathbb{I}_{\mathcal{C}} \rightsquigarrow$ assumption $\mathcal{C}$ bounded can be dropped
- $H_{u^*}(u^*) = \inf H_{u^*} \iff u^*$ minimizes $f + g$
  $\rightsquigarrow \quad D(u) := H_u(u) - \inf H_u \quad$ residual

HPGCG: Step-size

At $k \in \mathbb{N}$, **HPGCG** requires minimizing over H

$$v \mapsto H_{u^k}(v) := \langle \nabla f(u^k) - Pu^k, v \rangle + \tfrac{1}{2}\|v\|_P^2 + g(v).$$

Some **remarks**

- It requires $\frac{1}{2}\|\cdot\|_P^2 + g$ to be (strongly) **coercive**,

  if $g = \mathbb{I}_{\mathcal{C}} \rightsquigarrow$ assumption $\mathcal{C}$ bounded can be dropped
- $H_{u^*}(u^*) = \inf H_{u^*} \quad \Longleftrightarrow \quad u^*$ minimizes $f + g$

  $\rightsquigarrow \quad D(u) := H_u(u) - \inf H_u \quad$ residual
- **(Step-size)** If ∇f is *L-Lipschitz*, we choose

$$\theta_k := \min\left\{1, \frac{D(u^k) + \frac{1}{2}\|u^k - \widetilde{u}^k\|_P^2}{L\|u^k - \widetilde{u}^k\|^2}\right\}$$

$\sim$ *[B./Lorenz '08]*

HPGCG: Convergence result

Theorem *[B./Chenchene/Hosseini '23]*

Let $f, g \in \Gamma_0(H)$ and $P \succcurlyeq 0$ as before. Assume that

- $\frac{1}{2}\|\cdot\|_P^2 + g$ is (strongly) coercive
- ∇f is L-Lipschitz

HPGCG: Convergence result

Theorem *[B./Chenchene/Hosseini '23]*

Let $f, g \in \Gamma_0(H)$ and $P \succcurlyeq 0$ as before. Assume that

- $\frac{1}{2}\|\cdot\|_P^2 + g$ is (strongly) coercive
- ∇f is L-Lipschitz

Let $\{u^k\}_k$, $\{\widetilde{u}^k\}_k$ be generated by HPGCG with a specific step-size θ_k. Then

HPGCG: Convergence result

Theorem *[B./Chenchene/Hosseini '23]*

Let $f, g \in \Gamma_0(H)$ and $P \succcurlyeq 0$ as before. Assume that

- $\frac{1}{2}\|\cdot\|_P^2 + g$ is (strongly) coercive
- ∇f is L-Lipschitz

Let $\{u^k\}_k$, $\{\widetilde{u}^k\}_k$ be generated by HPGCG with a specific step-size θ_k. Then

1. $(f+g)(u^k) - \inf(f+g) = o(k^{-1/3})$ **monotonically**

HPGCG: Convergence result

Theorem *[B./Chenchene/Hosseini '23]*

Let $f, g \in \Gamma_0(H)$ and $P \succcurlyeq 0$ as before. Assume that

- $\frac{1}{2}\|\cdot\|_P^2 + g$ is (strongly) coercive
- ∇f is L-Lipschitz

Let $\{u^k\}_k$, $\{\widetilde{u}^k\}_k$ be generated by HPGCG with a specific step-size θ_k. Then

1. $(f+g)(u^k) - \inf(f+g) = o(k^{-1/3})$ **monotonically**
2. $\{u^k\}_k$, $\{\widetilde{u}^k\}_k$ are **bounded** and every weak cluster point is a minimizer of $f+g$

HPGCG: Convergence result

Theorem *[B./Chenchene/Hosseini '23]*

Let $f, g \in \Gamma_0(H)$ and $P \succcurlyeq 0$ as before. Assume that

- $\frac{1}{2}\|\cdot\|_P^2 + g$ is (strongly) coercive
- ∇f is L-Lipschitz

Let $\{u^k\}_k$, $\{\widetilde{u}^k\}_k$ be generated by HPGCG with a specific step-size θ_k. Then

1. $(f+g)(u^k) - \inf(f+g) = o(k^{-1/3})$ **monotonically**
2. $\{u^k\}_k$, $\{\widetilde{u}^k\}_k$ are **bounded** and every weak cluster point is a minimizer of $f+g$
3. $P^{1/2}u^k \rightharpoonup P^{1/2}u^*$ with u^* **minimizer** of $f+g$

Outline

TV parameter learning: Optimization

We consider $\|(A, v)\|_P^2 = \lambda \|A\|^2$ for some $\lambda > 0$

TV parameter learning: Optimization

We consider $\|(A, v)\|_P^2 = \lambda \|A\|^2$ for some $\lambda > 0$

The **semi-linearized** problem reads

$$\min_{v,A} \; -\frac{1}{N_t}\sum_{i=1}^{N_t} \langle \nabla(\operatorname{div} v_i^k + f_i), v_i\rangle + \frac{1}{N_t}\sum_{i=1}^{N_t} \mathrm{TV}(u_i^\dagger)\langle \bar{f}_i \otimes \bar{f}_i - \lambda A^k, A\rangle + \frac{\lambda}{2}\|A\|^2$$

$$\text{s.t.:} \quad A \succcurlyeq 0, \quad \|v_i\|_{\infty,2} \leq \bar{f}_i^* A \bar{f}_i \quad \forall i \in \{1, \ldots, N_t\}$$

TV parameter learning: Optimization

We consider $\|(A, v)\|_P^2 = \lambda \|A\|^2$ for some $\lambda > 0$

The **semi-linearized** problem reads

$$\min_{v,A} \; -\frac{1}{N_t}\sum_{i=1}^{N_t} \langle \nabla(\operatorname{div} v_i^k + f_i), v_i\rangle + \frac{1}{N_t}\sum_{i=1}^{N_t} \mathrm{TV}(u_i^\dagger)\langle \bar{f}_i \otimes \bar{f}_i - \lambda A^k, A\rangle + \frac{\lambda}{2}\|A\|^2$$

$$\text{s.t.:} \quad A \succcurlyeq 0, \quad \|v_i\|_{\infty,2} \leq \bar{f}_i^* A \bar{f}_i \quad \forall i \in \{1, \ldots, N_t\}$$

$\rightsquigarrow$ admits a **closed form solution**:

$$\widetilde{A}^k = P_{\succcurlyeq 0}\Big(A^k - \tfrac{1}{\lambda N_t}\sum_{i=1}^{N_t}\big(\mathrm{TV}(u_i^\dagger) - \mathrm{TV}(\operatorname{div} v_i^k + f_i)\big)\bar{f}_i \otimes \bar{f}_i\Big)$$

$$\widetilde{v}_i^k = \nabla(\operatorname{div} v_i^k + f_i)/|\nabla(\operatorname{div} v_i^k + f_i)|\bar{f}_i^* \widetilde{A}^k \bar{f}_i$$

HPGCG for TV parameter learning

- **HPGCG** turns into an efficient scheme:

HPGCG for TV parameter learning

- **HPGCG** turns into an efficient scheme:
  - Computations up to numerical precision (no inner loops),

HPGCG for TV parameter learning

- **HPGCG** turns into an efficient scheme:
  - Computations up to numerical precision (no inner loops),
  - Stronger **convergence**

$$A^k \to A^*, \quad \text{for some } A^* : \ \exists \boldsymbol{v}^* : (\boldsymbol{v}^*, A^*) \text{ is optimal}$$

HPGCG for TV parameter learning

- **HPGCG** turns into an efficient scheme:
  - Computations up to numerical precision (no inner loops),
  - Stronger **convergence**

$$A^k \to A^*, \quad \text{for some } A^* : \ \exists \boldsymbol{v}^* : (\boldsymbol{v}^*, A^*) \text{ is optimal}$$

Recall $\alpha : f \mapsto \bar{f}^* A^* \bar{f}$ is the predictor

HPGCG for TV parameter learning

- **HPGCG** turns into an efficient scheme:
  - Computations up to numerical precision (no inner loops),
  - Stronger **convergence**

$$A^k \to A^*, \quad \text{for some } A^* : \ \exists \boldsymbol{v}^* : (\boldsymbol{v}^*, A^*) \text{ is optimal}$$

Recall $\alpha : f \mapsto \bar{f}^* A^* \bar{f}$ is the predictor

- At present, we can only guarantee a $o(k^{-1/3})$ worst-case rate

Numerical experiments

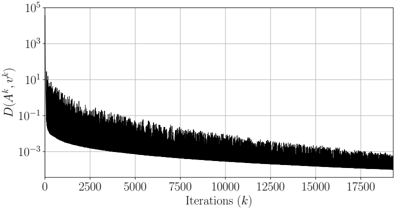
Dataset composed of 101440 16×16-patches from 1109 cartoon images. To each patch we apply a Gaussian noise

Numerical experiments

Dataset composed of 101440 16×16-patches from 1109 cartoon images. To each patch we apply a Gaussian noise

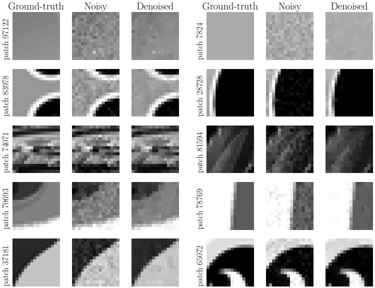
Implementation details

We set $\lambda = 50$, stopping criterion: $D(A^k, v^k) < 10^{-4}$ reached in $\sim 20k$ iterations



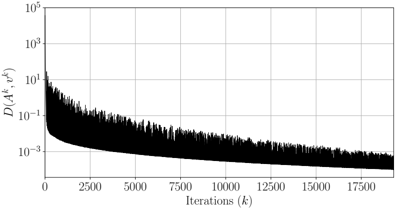
Numerical experiments

Dataset composed of 101440 16×16-patches from 1109 cartoon images. To each patch we apply a Gaussian noise



Implementation details

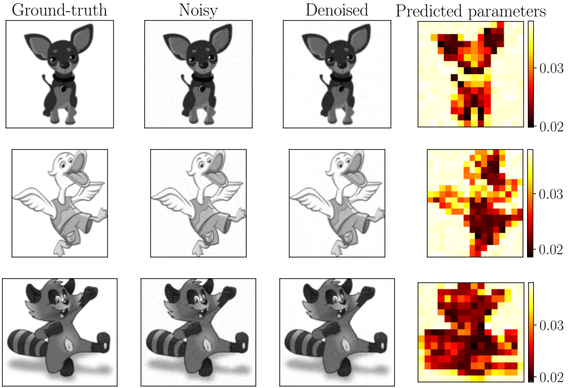
We set $\lambda = 50$, stopping criterion: $D(A^k, v^k) < 10^{-4}$ reached in $\sim 20k$ iterations



We can monitor the quality of the reconstruction online looking at $\operatorname{div} v_i^k + f_i$

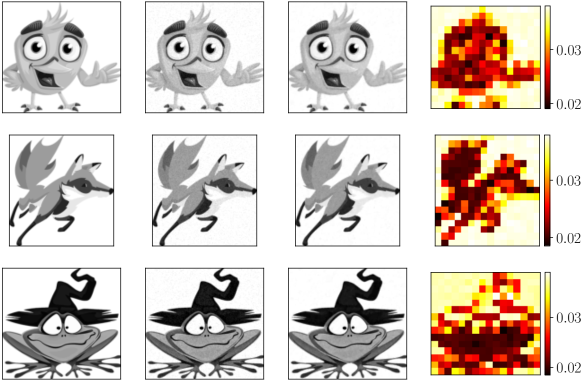
Numerical experiments

- The proposed model adaptively yields higher values for flatter regions (e.g., the backgrounds) and lower values for more complex parts of the images



Numerical experiments

- The proposed model adaptively yields higher values for flatter regions (e.g., the backgrounds) and lower values for more complex parts of the images



Numerical experiments

- On average, the model performs better than constant choices. As **performance metrics**, we consider

$$\mathrm{MSE}_\alpha := \frac{1}{N_t}\sum_{i=1}^{N_t}(\alpha_i^* - \alpha(f_i))^2, \quad \mathrm{MSE}_u := \frac{1}{N_t}\sum_{i=1}^{N_t}\|u_i^\dagger - u_i^{\alpha(f_i)}\|^2$$

Models	Quadratic	Constant $\alpha = \eta\ 10^{-4}$			
		$\eta = 1$	$\eta = 2.68$	$\eta = 7.20$	$\eta = 19.3$
MSE_α	**$3.39\ 10^{-4}$**	$18.56\ 10^{-4}$	$18.42\ 10^{-4}$	$18.08\ 10^{-4}$	$17.17\ 10^{-4}$
MSE_u	**0.1529**	0.6374	0.6312	0.6148	0.5729

Models	Constant $\alpha = \eta\ 10^{-3}$				
	$\eta = 5.18$	$\eta = 13.9$	$\eta = 27.13$	$\eta = 37.3$	$\eta = 100$
MSE_α	$14.88\ 10^{-4}$	$9.79\ 10^{-4}$	$4.95\ 10^{-4}$	$3.62\ 10^{-4}$	$41.08\ 10^{-4}$
MSE_u	0.4737	0.2917	0.1833	0.1777	0.4764

Outline

Conclusions and perspectives

We presented a new **Bregman**-learning framework

1. Investigate different models, *(NNs, ...)*

Conclusions and perspectives

We presented a new **Bregman**-learning framework

1. Investigate different models, *(NNs, ...)*
2. Learn proper, convex, lsc. regularizers

$$\alpha\,\mathrm{TV} \quad \text{learn } \alpha > 0 \quad \rightsquigarrow \quad g \quad \text{learn } g \in \Gamma_0(H)$$

Conclusions and perspectives

We presented a new **Bregman**-learning framework

1. Investigate different models, *(NNs, ...)*
2. Learn proper, convex, lsc. regularizers

$$\alpha\,\mathrm{TV} \quad \text{learn } \alpha > 0 \quad \rightsquigarrow \quad g \quad \text{learn } g \in \Gamma_0(H)$$

3. Better understand the loss function (connection to Weak Optimal Transport)

Conclusions and perspectives

We presented a new **Bregman**-learning framework

1. Investigate different models, *(NNs, ...)*
2. Learn proper, convex, lsc. regularizers

$$\alpha\,\mathrm{TV} \quad \text{learn } \alpha > 0 \quad \rightsquigarrow \quad g \quad \text{learn } g \in \Gamma_0(H)$$

3. Better understand the loss function (connection to Weak Optimal Transport)
4. Devise **stochastic** hybrid methods for training

Conclusions and perspectives

We presented a new **Bregman**-learning framework

1. Investigate different models, *(NNs, ...)*
2. Learn proper, convex, lsc. regularizers

$$\alpha\,\mathrm{TV} \quad \text{learn } \alpha > 0 \quad \rightsquigarrow \quad g \quad \text{learn } g \in \Gamma_0(H)$$

3. Better understand the loss function (connection to Weak Optimal Transport)
4. Devise **stochastic** hybrid methods for training

E. Chenchene, A. Hosseini and K. Bredies.
A hybrid proximal generalized conditional gradient method and application to total variation parameter learning.
In: 2023 European Control Conference (ECC), pp. 322–327, 2023.

This work has received funding from the European Union's Framework Programme for Research and Innovation Horizon 2020 (2014–2020) under the Marie Skłodowska-Curie Grant Agreement No. 861137