Chapter 4

Primal-Dual Algorithms

Convex Optimization for Computer Vision

SS 2016

Michael Moeller
Thomas Möllenhoff
Emanuel Laude
Computer Vision Group
Department of Computer Science
TU München

Recalling gradient methods

Michael Moeller
Thomas Möllenhoff
Emanuel Laude

Structured problems

1 Gradient descent:

$$\min_u E(u)$$

for $E : \mathbb{R}^n \to \mathbb{R}$ L-smooth: $\mathcal{O}(1/k)$,

2 Subgradient descent:

$$\min_u E(u)$$

$E : \mathbb{R}^n \to \mathbb{R}$ Lipschitz continuous, stepsizes$\to$ 0: $\mathcal{O}(1/\sqrt{k})$.

3 Proximal gradient:

$$\min_u F(u) + G(u)$$

$F : \mathbb{R}^n \to \mathbb{R}$ L-smooth, $G : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ simple: $\mathcal{O}(1/k)$.

3* Gradient projection: Special case of prox. grad. for $G = \iota_C$

Strong convexity: Linear convergence $\mathcal{O}(c^k), c < 1$, of **1** and **3**.

Michael Moeller
Thomas Möllenhoff
Emanuel Laude

Structured problems

How would you solve

$$\min_u \frac{1}{2}\|u - f\|_1 + \alpha\|Ku\|_2^2$$

$\rightarrow$ Proximal gradient

How would you solve

$$\min_u \frac{1}{2}\|u - f\|^2 + \alpha\|Ku\|_2^2$$

Gradient Descent (although there are better ways).

How would you solve

$$\min_u \frac{1}{2}\|u - f\|^2 + \alpha\|Ku\|_1$$

$\rightarrow$ Derive dual problem, apply gradient projection

Michael Moeller
Thomas Möllenhoff
Emanuel Laude

Structured problems

How would you solve

$$\min_u \frac{1}{2}\|u - f\|_1 + \alpha\|Ku\|_1$$

$\rightarrow$ Subgradient descent

Is this really the best we can do for such a problem? No!

Very important class of algorithms we have not considered yet!

Applicable to

$$\min_u G(u) + F(Ku)$$

with $G : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\infty\}$ simple[1], $F : \mathbb{R}^m \rightarrow \mathbb{R} \cup \{\infty\}$ simple, $K : \mathbb{R}^n \rightarrow \mathbb{R}^m$ linear.

[1] simple = easy to evaluate proximity operator

Primal-dual hybrid gradient method

Reminders

The convex-conjugate of E is

$$E^*(p) = \sup_u \langle u, p \rangle - E(u)$$

Fact 1: For a proper, closed, convex E it holds that

$$E = E^{**}$$

Trick to optimize the TV functional: Use

$$\|Ku\|_{2,1} = (\|\cdot\|_{2,1})^{**}(Ku) = \sup_{\|p\|_{2,\infty} \leq 1} \langle Ku, p \rangle$$

Reminders

Proximal operator

$$\text{prox}_E(v) := \underset{u}{\text{argmin}}\, E(u) + \frac{1}{2}\|u - v\|^2$$

Moreau decomposition:

$$v = \text{prox}_E(v) + \text{prox}_{E^*}(v)$$

Fact 2: If E is simple, E^* is simple, too!

Fact 3: The operator $\text{prox}_{\tau E}$ has the interpretation of an implicit gradient descent step (even for nondifferentiable functions).

Opposed to explicit (sub-)gradient descent it guarantees the reduction of E for any τ

Primal-dual formulation

Let us consider

$$\min_u G(u) + F(Ku)$$

with G and F being simple.

Let us try what helped us in our first computation for deriving duality: Use "fact 1": $F = F^{**}$

$$\min_u G(u) + F(Ku) = \min_u \sup_p G(u) + \langle Ku, p \rangle - F^*(p)$$

Based on "fact 2", prox_{F^*} is easy to evaluate.

Let's try to alternate between an implicit gradient descent on u and an implicit gradient ascent in p!

Michael Moeller
Thomas Möllenhoff
Emanuel Laude

Primal-dual formulation

Define

$$PD(u,p) := G(u) + \langle Ku, p\rangle - F^*(p)$$

and try

$$\begin{aligned} p^{k+1} &= \text{prox}_{-\sigma PD(u^k,\cdot)}(p^k), \\ u^{k+1} &= \text{prox}_{\tau PD(\cdot,p^{k+1})}(u^k), \end{aligned}$$

One finds

$$\begin{aligned} p^{k+1} &= \text{prox}_{-\sigma PD(u^k,\cdot)}(p^k), \\ &= \operatorname*{argmin}_p \frac{1}{2}\|p - p^k\|^2 + \sigma F^*(p) - \sigma\langle Ku^k, p\rangle \\ &= \operatorname*{argmin}_p \frac{1}{2}\|p - p^k - \sigma Ku^k\|^2 + \sigma F^*(p) \\ &= \text{prox}_{\sigma F^*}(p^k + \sigma Ku^k) \end{aligned}$$

Michael Moeller
Thomas Möllenhoff
Emanuel Laude

Primal-dual formulation

Define

$$PD(u,p) := G(u) + \langle Ku, p\rangle - F^*(p)$$

and try

$$\begin{aligned} p^{k+1} &= \text{prox}_{\sigma F^*}(p^k + \sigma K u^k), \\ u^{k+1} &= \text{prox}_{\tau PD(\cdot, p^{k+1})}(u^k), \end{aligned}$$

One finds

$$\begin{aligned} u^{k+1} &= \text{prox}_{\tau PD(\cdot, p^{k+1})}(u^k), \\ &= \operatorname*{argmin}_u \frac{1}{2}\|u - u^k\|^2 + G(u) + \langle Ku, p^{k+1}\rangle \\ &= \operatorname*{argmin}_u \frac{1}{2}\|u - u^k + \tau K^* p^{k+1}\|^2 + \tau G(u) \\ &= \text{prox}_{\tau G}(u^k - \tau K^* p^{k+1}) \end{aligned}$$

Primal-dual hybrid gradient method

We found

$$
\begin{aligned}
p^{k+1} &= \text{prox}_{\sigma F^*}(p^k + \sigma K u^k), \\
u^{k+1} &= \text{prox}_{\tau G}(u^k - \tau K^* p^{k+1}).
\end{aligned}
$$

One should make one (currently unintuitive) modification:

$$
\begin{aligned}
p^{k+1} &= \text{prox}_{\sigma F^*}(p^k + \sigma K \bar{u}^k), \\
u^{k+1} &= \text{prox}_{\tau G}(u^k - \tau K^* p^{k+1}), \qquad \text{(PDHG)} \\
\bar{u}^{k+1} &= 2u^{k+1} - u^k.
\end{aligned}
$$

We will understand this modification very well in about 2 weeks!

Our goal: Prove that the **Primal-Dual Hybrid Gradient Method**[2] (PDHG) converges!

[2] Pock, Cremers, Bischof, Chambolle '08, Esser, Zhang, Chan '09, Chambolle, Pock '10

Michael Moeller
Thomas Möllenhoff
Emanuel Laude

Saddle points

We wrote

$$\min_u G(u) + F(Ku) = \min_u \sup_p PD(u, p)$$

where

$$PD(u, p) = G(u) + \langle Ku, p\rangle - F^*(p)$$

for proper, closed, convex G and F. We assume that a minimizer $\tilde{u}$ exists. Is the "sup" attained for some $\tilde{p}$, too?

Definition

Let we call $(\tilde{u}, \tilde{p})$ a *saddle-point* of $PD : \mathbb{R}^n \times \mathbb{R}^m \to \bar{\mathbb{R}}$, if

$$PD(\tilde{u}, p) \leq PD(\tilde{u}, \tilde{p}) \leq PD(u, \tilde{p})$$

holds for all $u \in \mathbb{R}^n$, $p \in \mathbb{R}^m$.

Michael Moeller
Thomas Möllenhoff
Emanuel Laude

Existing saddle-point

1. Let there exist $\tilde{u} \in \text{argmin}_u \, G(u) + F(Ku)$,
2. Let there exist a $u \in \text{ri}(\text{dom}(G))$ such that $Ku \in \text{ri}(\text{dom}(F))$.

Then (according to the sum rule)

$$0 \in \partial G(\tilde{u}) + K^* \partial F(K\tilde{u}).$$

In particular, F is subdifferentiable at $K\tilde{u}$ and we know that

$$\sup_p \langle K\tilde{u}, p \rangle - F^*(p) = F^{**}(K\tilde{u}) = \langle K\tilde{u}, \tilde{p} \rangle - F^*(\tilde{p})$$

for $\tilde{p} \in \partial F(K\tilde{u})$ according to the Fenchel-Young inequality.

$\rightarrow$ Under 1. and 2., a saddle point of *PD* exists!

Partial primal-dual gap

Definition

Given two compact sets B_1 and B_2, we call

$$\mathcal{G}_{B_1 \times B_2}(u,p) = \max_{p' \in B_2} PD(u,p') - \min_{u' \in B_1} PD(u',p)$$

the *partial primal-dual gap*.

Properties of the partial primal-dual gap

Let $(\tilde{u}, \tilde{p})$ be a saddle point of the min-max problem

$$\min_u \max_p G(u) + \langle Ku, p \rangle - F^*(p),$$

and let $(\tilde{u}, \tilde{p}) \in B_1 \times B_2$. Then

$$\mathcal{G}_{B_1 \times B_2}(u,p) \geq 0 \qquad \forall (u,p) \in \mathbb{R}^n \times \mathbb{R}^m$$

$$\mathcal{G}_{B_1 \times B_2}(u,p) = 0 \;\Leftrightarrow\; (u,p) \text{ is a saddle point.}$$

Michael Moeller
Thomas Möllenhoff
Emanuel Laude

PDHG convergence theorem

Theorem (Chambolle-Pock '10)

In addition to the previous assumptions, let $L = \|K\|$, let $\tau\sigma L^2 < 1$, and let $(p^n, u^n, \bar{u}^n)$ be the iterates of (PDHG) for an arbitrary starting point u^0, p^0 and $\bar{u}^0 = u^0$.

Then (u^n, p^n) converge to a saddle-point (u^*, p^*) of PD. For $u_N = (\sum_{k=1}^N u^k)/N$, $p_N = (\sum_{k=1}^N p^k)/N$, and any compact $B_1 \times B_2 \subset \mathbb{R}^n \times \mathbb{R}^m$, it holds that

$$\mathcal{G}_{B_1 \times B_2}(u_N, p_N) \leq \frac{D(B_1, B_2)}{N},$$

where

$$D(B_1, B_2) = \max_{(u,p) \in B_1 \times B_2} \frac{\|u - u^0\|^2}{2\tau} + \frac{\|p - p^0\|^2}{2\sigma},$$

and (u_N, p_N) also converge to (u^*, p^*).

Proof: Board.

Michael Moeller
Thomas Möllenhoff
Emanuel Laude

PDHG

The primal-dual hybrid gradient method

$$p^{k+1} = \text{prox}_{\sigma F^*}(p^k + \sigma K \bar{u}^k),$$

$$u^{k+1} = \text{prox}_{\tau G}(u^k - \tau K^* p^{k+1}), \qquad \text{(PDHG)}$$

$$\bar{u}^{k+1} = u^{k+1} + (u^{k+1} - u^k).$$

ROF Denoising

$$\min P(u) = \min_u \frac{1}{2}\|u - f\|^2 + \alpha\|Ku\|_{2,1}$$

with K being a discretization of the multichannel gradient operator.

ROF Denoising

We write

$$\min_u P(u) = \min_u \max_p \frac{1}{2}\|u - f\|^2 + \langle Ku, p\rangle - \iota_{\|\cdot\|_{2,\infty}\leq\alpha}(p).$$

The (PDHG) updates are

$$\begin{aligned} p^{k+1} &= \text{prox}_{\sigma F^*}(p^k + \sigma K \bar{u}^k) \\ u^{k+1} &= \text{prox}_{\tau G}(u^k - \tau K^* p^{k+1}), \\ \bar{u}^{k+1} &= 2u^{k+1} - u^k. \end{aligned}$$

which in this case amounts to

$$\begin{aligned} p^{k+1} &= \operatorname*{argmin}_p \frac{1}{2}\|p - (p^k + \sigma K \bar{u}^k)\|^2 + \sigma \iota_{\|\cdot\|_{2,\infty}\leq\alpha}(p), \\ u^{k+1} &= \operatorname*{argmin}_u \frac{1}{2}\|u - (u^k - \tau K^* p^{k+1})\|^2 + \frac{\tau}{2}\|u - f\|^2 \\ &= \frac{u^k - \tau K^* p^{k+1} + \tau f}{1+\tau} \\ \bar{u}^{k+1} &= 2u^{k+1} - u^k. \end{aligned}$$

TV-L^1 Denoising

$$\min P(u) = \min_u \|u - f\|_1 + \alpha \|Ku\|_{2,1}$$

with K being a discretization of the multichannel gradient operator.

TV-L^1 Denoising

We write

$$\min_u P(u) = \min_u \max_p \frac{1}{2}\|u - f\|_1 + \langle Ku, p\rangle - \iota_{\|\cdot\|_{2,\infty}\leq\alpha}(p).$$

The (PDHG) updates are

$$\begin{aligned} p^{k+1} &= \text{prox}_{\sigma F^*}(p^k + \sigma K\bar{u}^k) \\ u^{k+1} &= \text{prox}_{\tau G}(u^k - \tau K^* p^{k+1}), \\ \bar{u}^{k+1} &= 2u^{k+1} - u^k. \end{aligned}$$

which in this case amounts to

An exercise! :-)

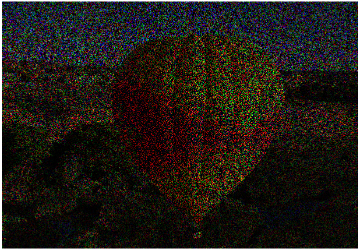
Michael Moeller
Thomas Möllenhoff
Emanuel Laude

TV-Inpainting

$$\min P(u) = \min_u \iota_{f_{|I}}(u) + \alpha \|Ku\|_{2,1}$$

with K being a discretization of the color gradient operator, and

$$\iota_{f_{|I}}(u) = \begin{cases} 0 & \text{if } u_i = f_i \text{ for all } i \in I, \\ \infty & \text{otherwise.} \end{cases} .$$



Michael Moeller
Thomas Möllenhoff
Emanuel Laude

TV-Inpainting

We write

$$\min_u P(u) = \min_u \max_p \iota_{f_{|I}}(u) + \langle Ku, p\rangle - \iota_{\|\cdot\|_{2,\infty}\leq\alpha}(p).$$

The (PDHG) updates are

$$p^{k+1} = \text{prox}_{\sigma F^*}(p^k + \sigma K\bar{u}^k)$$
$$u^{k+1} = \text{prox}_{\tau G}(u^k - \tau K^* p^{k+1}),$$
$$\Rightarrow u_i^{k+1} = \begin{cases} f_i & \text{if } i \in I, \\ (u^k - \tau K^* p^{k+1})_i & \text{otherwise.} \end{cases}$$
$$\bar{u}^{k+1} = 2u^{k+1} - u^k.$$

Michael Moeller
Thomas Möllenhoff
Emanuel Laude

TV-Deblurring

$$\min P(u) = \min_u \frac{1}{2}\|Au - f\|^2 + \alpha\|Ku\|_{2,1}$$

with K being a discretization of the multichannel gradient operator, A being a convolution with a blur kernel.

Michael Moeller
Thomas Möllenhoff
Emanuel Laude

TV-Deblurring - Option 1

We write

$$\min_u P(u) = \min_u \max_p \frac{1}{2}\|Au - f\|^2 + \langle Ku, p\rangle - \iota_{\|\cdot\|_{2,\infty}\leq\alpha}(p).$$

The (PDHG) updates are

$$\begin{aligned} p^{k+1} &= \text{prox}_{\sigma F^*}(p^k + \sigma K\bar{u}^k) \\ u^{k+1} &= \text{prox}_{\tau G}(u^k - \tau K^* p^{k+1}), \\ \bar{u}^{k+1} &= 2u^{k+1} - u^k. \end{aligned}$$

which in this case amounts to

$$\begin{aligned} p^{k+1} &= \arg\min_p \frac{1}{2}\|p - (p^k + \sigma K\bar{u}^k)\|^2 + \sigma\iota_{\|\cdot\|_{2,\infty}\leq\alpha}(p), \\ u^{k+1} &= \arg\min_u \frac{1}{2}\|u - (u^k - \tau K^* p^{k+1})\|^2 + \frac{\tau}{2}\|Au - f\|^2 \\ &= (I + \tau A^* A)^{-1}(u^k - \tau K^* p^{k+1} + \tau f) \\ \bar{u}^{k+1} &= 2u^{k+1} - u^k. \end{aligned}$$

Michael Moeller
Thomas Möllenhoff
Emanuel Laude

TV-Deblurring - Option 2

We write

$$
\begin{aligned}
&\min_u P(u)\\
&= \min_u \max_{p,q} \langle Au - f, q\rangle - \frac{1}{2}\|q\|^2 + \langle Ku, p\rangle - \iota_{\|\cdot\|_{2,\infty}\leq\alpha}(p)\\
&= \min_u \max_{p,q} \left\langle \begin{pmatrix} A \\ K \end{pmatrix} u, \begin{pmatrix} q \\ p \end{pmatrix} \right\rangle - \langle f, q\rangle - \frac{1}{2}\|q\|^2 - \iota_{\|\cdot\|_{2,\infty}\leq\alpha}(p)
\end{aligned}
$$

Now we have

$$
\begin{aligned}
F^*(p,q) &= \langle f, q\rangle + \frac{1}{2}\|q\|^2 + \iota_{\|\cdot\|_{2,\infty}\leq\alpha}(p)\\
G(u) &= 0\\
\tilde{K} &= \begin{pmatrix} A \\ K \end{pmatrix}
\end{aligned}
$$

TV-Deblurring - Option 2

The (PDHG) updates are

$$
\begin{aligned}
q^{k+1} &= \operatorname*{argmin}_q \frac{1}{2}\|q - (q^k + \sigma A\bar{u}^k)\|^2 + \sigma\langle f, q\rangle + \frac{\sigma}{2}\|q\|^2,\\
p^{k+1} &= \operatorname*{argmin}_p \frac{1}{2}\|p - (p^k + \sigma K\bar{u}^k)\|^2 + \sigma\iota_{\|\cdot\|_{2,\infty}\leq\alpha}(p),\\
u^{k+1} &= u^k - \tau K^* p^{k+1} - \tau A^* q^{k+1}\\
\bar{u}^{k+1} &= 2u^{k+1} - u^k.
\end{aligned}
$$

Michael Moeller
Thomas Möllenhoff
Emanuel Laude

TV-Zooming

$$\min P(u) = \min_u \frac{1}{2}\|Au - f\|^2 + \alpha\|Ku\|_{2,1}$$

with K being a discretization of the multichannel gradient operator, $A = DB$, with B being a convolution with a blur kernel, and D being a downsampling, e.g. a matrix

$$D = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & \ldots & \ldots \\ 0 & 0 & 1 & 0 & 0 & \ldots & \ldots \\ 0 & 0 & 0 & 0 & 1 & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \end{pmatrix}$$

PDHG implementation: Option 2 from the previous example.

Michael Moeller
Thomas Möllenhoff
Emanuel Laude

TV-Zooming

Input data

Nearest neighbor

TV Zooming

Image Segmentation

$$\min P(u) = \min_u \iota_\Delta(u) + \iota_{\geq 0}(u) + \langle u, f \rangle + \alpha \|Ku\|_{2,1}$$

where $K : \mathbb{R}^{n \times m \times c} \to \mathbb{R}^{nmc \times 2}$ being a discretization of the multichannel gradient operator, and

$$\iota_\Delta(u) = \begin{cases} 0 & \text{if } \sum_k u_{i,j,k} = 1, \ \forall (i,j) \\ \infty & \text{else.} \end{cases}$$

$$\iota_{\geq 0}(u) = \begin{cases} 0 & \text{if } u_{i,j,k} \geq 0, \ \forall (i,j,k) \\ \infty & \text{else.} \end{cases}$$

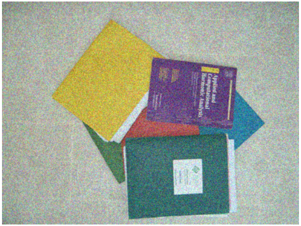
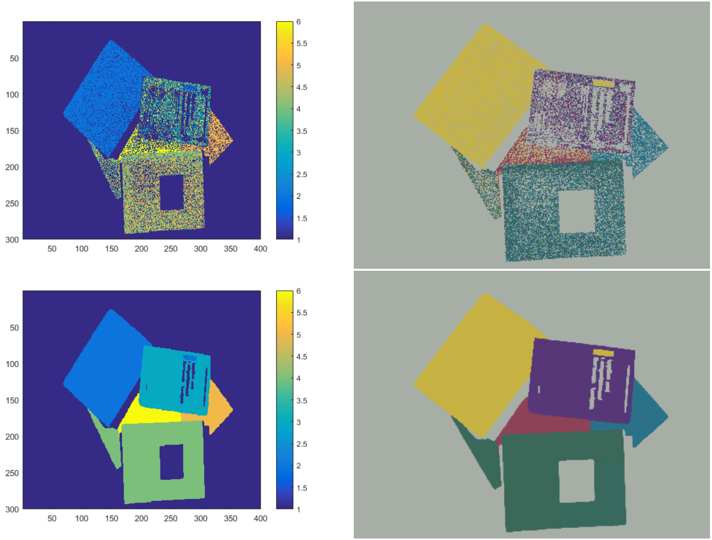


Image Segmentation



Upper row: data term minimization (=kmeans assignment), lower row: variational method

Michael Moeller
Thomas Möllenhoff
Emanuel Laude

Image Segmentation

Option 1: We solve

$$\min_{u} \max_{p} \iota_{\Delta}(u) + \iota_{\geq 0}(u) + \langle u, f \rangle + \langle Ku, p \rangle - \iota_{\|\cdot\|_{2,\infty} \leq \alpha}(p).$$

$\rightarrow$ Primal proximal operator: Projection onto unit simplex.

Option 2: We solve

$$\min_{u} \max_{p.q} \langle Su - 1, q \rangle + \iota_{\geq 0}(u) + \langle u, f \rangle + \langle Ku, p \rangle - \iota_{\|\cdot\|_{2,\infty} \leq \alpha}(p).$$

where $(Su)_{i,j} = \sum_k u_{i,j}$.

$\rightarrow$ Very simple proximal operators, but additional variable.

Michael Moeller
Thomas Möllenhoff
Emanuel Laude

Excursus: Further regularizations

and

noncovex minimization through sequences of convex problems.

Beyond total variation regularization

We used $J(u) = \|Ku\|_{2,1}$ in most applications.

Pros:

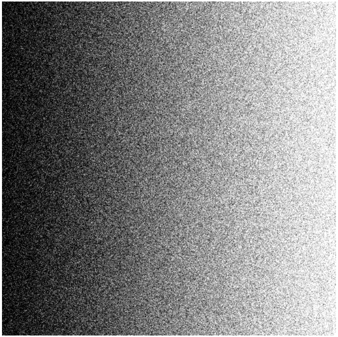
- Preserves discontinuities/jumps
- Easy to use and fast to implement
- Good quality on images that are well approximated by piecewise constant images

Cons:

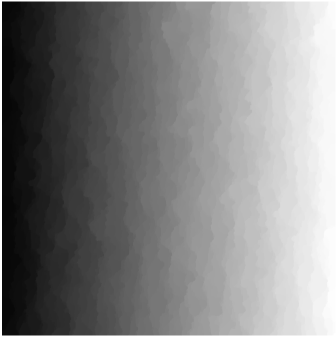
- Staircasing
- Not necessarily a realistic image prior

Michael Moeller
Thomas Möllenhoff
Emanuel Laude

Total generalized variation[3]



Input image



TV denoising

Can one avoid this ”staircasing” effect while preserving the ability to reconstruct jumps?

[3] *Total Generlaized Variation*, Bredis, Kunisch, Pock, SIAM Imag. Sci., 2010

Total generalized variation

Regularization

$$
\begin{aligned}
TGV(u) &= \min_{Ku = z_1 + z_2} \|z_1\|_{2,1} + \beta \|\tilde{K} z_2\|_{2,1} \\
&= \min_{z} \|Ku - z\|_{2,1} + \beta \|\tilde{K} z\|_{2,1}
\end{aligned}
$$

with $\tilde{K} : \mathbb{R}^{nm \times 2c} \to \mathbb{R}^{nm \times 4c}$ being a discretization of the derivative of a vector field.

Idea: Optimally divide into a penalty of the first derivative (TV) and the Hessian.

TGV denoising

Consider the TGV denoising problem

$$\min_{u,z} \frac{1}{2}\|u - f\|_2^2 + \alpha \left(\|Ku - z\|_{2,1} + \beta \|\tilde{K} z\|_{2,1} \right)$$

How can we minimize such a function (with PDHG)?

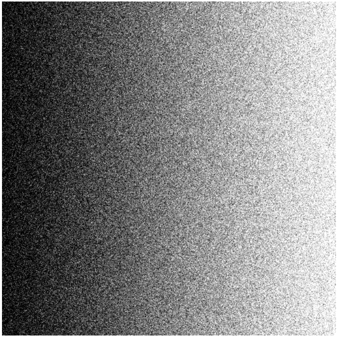
$$\min_{u,z} \frac{1}{2}\|u - f\|_2^2 + \alpha \left\| \begin{pmatrix} Ku - z \\ \beta \tilde{K} z \end{pmatrix} \right\|_{2,1}$$

$$= \min_{u,z} \frac{1}{2}\|u - f\|_2^2 + \alpha \left\| \underbrace{\begin{pmatrix} K & -I \\ 0 & \tilde{K} \end{pmatrix}}_{=:D} \underbrace{\begin{pmatrix} u \\ z \end{pmatrix}}_{=:v} \right\|_{2,1}$$

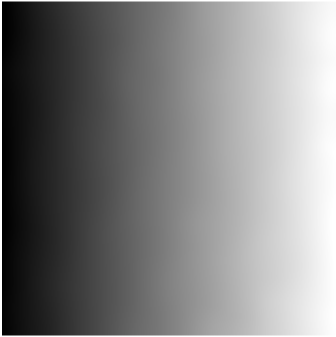
$$= G(v) + F(Dv)$$

Michael Moeller
Thomas Möllenhoff
Emanuel Laude

TGV denoising



Input image



TGV denoising

Michael Moeller
Thomas Möllenhoff
Emanuel Laude

Does TGV denoising solve everything?

Input image

TGV denoising

Textures are neither well reconstructed by TV nor by TGV.

Michael Moeller
Thomas Möllenhoff
Emanuel Laude

Problems of TV with textures

Input image

TV denoising

Textures are neither well reconstructed by TV nor by TGV.

Michael Moeller
Thomas Möllenhoff
Emanuel Laude

Nonlocal TV

Input image

NLTV denoising

Idea: Use self-similarity in the image!

Nonlocal TV [4] based on NL-Means[5]

Construct a similarity graph within an image:

$$w_{i,j} = \exp\left(-\frac{\|\text{patch around pixel i} - \text{patch around pixel j}\|^2}{2\sigma}\right)$$

Anisotropic NLTV regularization:

$$J(u) = \sum_{i,j} \sqrt{w_{i,j}}\, |u_i - u_j|$$

Illustration of a similar technique:

`http://www.cs.tut.fi/~foi/GCF-BM3D/`

Problem: Construction and size of $W = (w_{i,j}) \in \mathbb{R}^{\#pixels \times \#pixels}$

[4] *Nonlocal operators with applications to img. proc.*, Gilboa, Osher, 2007

[5] *A non-local algorithm for image denoising*, Buades, Coll, Morel, 2005

Michael Moeller
Thomas Möllenhoff
Emanuel Laude

Dictionary learning

Technique that is also based on self-similarity but shifts the problem of constructing a full similarity matrix W:

Dictionary learning: Any patch can be represented as a sparse linear combination of a few dictionary patches.

$$J(u) = \min_{D,\alpha} \|Au - D\alpha\|^2 + \gamma\|\alpha\|_1,$$

where $u \in \mathbb{R}^{n,m,c}$ is an image, $A : \mathbb{R}^{n \times m \times c} \to \mathbb{R}^{p_1 \; p_2 \; c \times T}$ extracts T color patches of size $p_1 \times p_2$ from the image, and $D \in \mathbb{R}^{p_1 \; p_2 \; c \times t}$ is a matrix or dictionary of t different patches.

The coefficients $\alpha \in \mathbb{R}^{t \times T}$ represent the patches Au.

Usually $p_1 \cdot p_2 \cdot c < t << T$.

Dictionary learning

Problems of the form

$$\min_{u,D,\alpha} \left(E(u,D,\alpha) := G(u) + J(u,D,\alpha) \right),$$

for

$$J(u,D,\alpha) = \min_{D,\alpha} \|Au - D\alpha\|^2 + \gamma\|\alpha\|_1,$$

are not convex, but at least *biconvex*.

Common strategy: Alternating minimization!

$$\begin{aligned}(u^{k+1},\alpha^{k+1}) &= \operatorname*{argmin}_{u,\alpha} E(u,D^k,\alpha)\\ D^{k+1} &= \operatorname*{argmin}_{D} E(u^{k+1},D,\alpha^{k+1})\end{aligned}$$

Dictionary learning: See e.g. *K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation*, Aharon, Elad, Bruckstein, 2006, *Sparse representation for color image restoration*, Mairal, Elad, Sapiro, 2008.

Michael Moeller
Thomas Möllenhoff
Emanuel Laude

Convex optimization is important even for nonconvex optimization!

Landweber method

$$\min_u G(u) + F(u),$$

G is nonconvex, but differentiable, F is simple: Proximal gradient method!

Iteratively-regularized Gauss-Newton / iterative linearization

$$\min_u \frac{1}{2}\|T(u) - f\|^2 + F(Ku),$$

where T is a nonlinear operator:

$$T(u) \approx T(u^k) + \nabla T(u^k) \cdot (u - u^k)$$

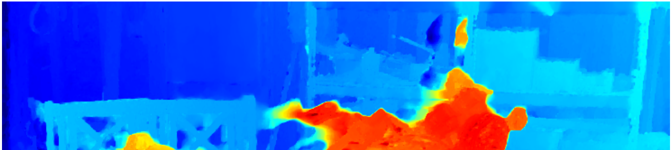
which results in

$$u^{k+1} = \operatorname*{argmin}_u \frac{1}{2}\|T(u^k) + \nabla T(u^k) \cdot (u - u^k) - f\|^2 + F(Ku).$$

Michael Moeller
Thomas Möllenhoff
Emanuel Laude

Stereo Matching

$$\int_\Omega |I_1(x+v(x),y) - I_2(x,y)|^2 \, dx \, dy + \alpha R(v)$$



Back to PDHG!

One strongly convex function

We have seen: PDHG works very well on problems of the form

$$\min G(u) + F(Ku),$$

for which F and G are simple.

We get the ergodic convergence of the partial primal-dual gap to zero with rate $\mathcal{O}(1/k)$.

What if our problem is ”more friendly”? E.g. what if G or F or both are strongly convex?

Modifications of PDHG

Michael Moeller
Thomas Möllenhoff
Emanuel Laude

One strongly convex function

Case 1: Either G or F^* is strongly convex

Adaptive stepsizes:

$$\begin{aligned}
p^{k+1} &= \text{prox}_{\sigma_k F^*}(p^k + \sigma_k K \bar{u}^k), \\
u^{k+1} &= \text{prox}_{\tau_k G}(u^k - \tau_k K^* p^{k+1}), \\
\theta_k &= \frac{1}{\sqrt{1 + 2\gamma\tau_k}}, \\
\tau_{k+1} &= \theta_k \tau_k, \quad \sigma_{k+1} = \sigma_k / \theta_k \\
\bar{u}^{k+1} &= u^{k+1} + \theta_k (u^{k+1} - u^k).
\end{aligned} \tag{PDHG2}$$

for $\tau_0 \sigma_0 \leq \|K\|^2$, and G being γ-strongly convex.

Improved convergence because we gain $\frac{\gamma}{2}\|u - u^{n+1}\|^2$ in the subgradient inequality.

One strongly convex function

Doing an estimate similar to the one of (PDHG) one obtains

$$\|\tilde{u} - u^N\|^2 \leq \tau_N^2 \left(\frac{\|\tilde{u} - u^0\|^2}{\tau_0^2} + \|K\|^2 \|\tilde{p} - p^0\|^2 \right)$$

How fast does τ_N decay?

$$\lim_{N \to \infty} N \tau_N = \frac{1}{\gamma}$$

Theorem about (PDHG2), strongly convex G, Chambolle, Pock '10

For any $\epsilon > 0$ there exists an N_0 such that for any $N \geq N_0$:

$$\|\tilde{u} - u^N\|^2 \leq \frac{1+\epsilon}{\gamma^2 N^2} \left(\frac{\|\tilde{u} - u^0\|^2}{\tau_0^2} + \|K\|^2 \|\tilde{p} - p^0\|^2 \right)$$

One strongly convex function

Discussion of the convergence orders:

- Didn't the gradient methods obtain linear convergence on strongly convex energies?
- Yes, but we additionally needed a part of the energy to be L-smooth!
- Note that L-smoothness of the primal corresponds to $1/L$-strong convexity of the dual!
- So the (PDHG2) is faster (asymptotically $\mathcal{O}(1/N^2)$) than proximal gradient with $\mathcal{O}(1/N)$?
- For the proximal gradient we discussed, yes! But there are adaptive proximal gradient schemes that reach $\mathcal{O}(1/N^2)$!
- Can we massage (PDHG) to have linear convergence if both, G and F^*, are strongly convex?

Two strongly convex functions

Consider

$$
\begin{aligned}
p^{k+1} &= \text{prox}_{\sigma F^*}(p^k + \sigma K \bar{u}^k), \\
u^{k+1} &= \text{prox}_{\tau G}(u^k - \tau K^* p^{k+1}), \\
\bar{u}^{k+1} &= u^{k+1} + \theta(u^{k+1} - u^k).
\end{aligned} \qquad \text{(PDHG2)}
$$

Chambolle, Pock '10

For $\mu \leq 2\sqrt{\gamma\delta}/\|K\|$, $\tau = \mu/(2\gamma)$, $\sigma = \mu/(2\delta)$, $\theta \in [1/(1+\mu), 1]$, G being γ-strongly convex and F^* being δ-strongly convex, there exists $\omega < 1$, such that the iterates of (PDHG2) meet

$$\gamma\|u^N - \tilde{u}\|^2 + (1-\omega)\delta\|p^N - \tilde{p}\|^2 \leq \omega^N(\gamma\|u^0 - \tilde{u}\|^2 + \delta\|p^0 - \tilde{p}\|^2).$$

$\rightarrow$ Linear convergence!

Understanding the (PDHG) structure

In many image processing and computer vision tasks, we will have to live with the simplest (PDHG) version

$$\begin{aligned} u^{k+1} &= \text{prox}_{\tau G}(u^k - \tau K^* p^k), \\ \bar{u}^{k+1} &= u^{k+1} + (u^{k+1} - u^k), \\ p^{k+1} &= \text{prox}_{\sigma F^*}(p^k + \sigma K \bar{u}^{k+1}), \end{aligned} \tag{PDHG}$$

and convergence of the partial primal-dual gap with $\mathcal{O}(1/k)$.

One thing that still remains somewhat unclear is why the extrapolation in $\bar{u}^{k+1}$ seems to stabilize the algorithm.

$\rightarrow$ *Exciting computation on the board!*

Michael Moeller
Thomas Möllenhoff
Emanuel Laude