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

Michael Moeller
Thomas Möllenhoff
Emanuel Laude

Michael Moeller
Thomas Möllenhoff
Emanuel Laude

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

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

Michael Moeller
Thomas Möllenhoff
Emanuel Laude

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

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

Primal-dual formulation

Define

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

and try

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

One finds

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

Michael Moeller
Thomas Möllenhoff
Emanuel Laude

Primal-dual hybrid gradient method

We found

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

One should make one (currently unintuitive) modification:

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

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

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!

Michael Moeller
Thomas Möllenhoff
Emanuel Laude

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.

PDHG

The primal-dual hybrid gradient method

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

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

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

Michael Moeller
Thomas Möllenhoff
Emanuel Laude

ROF Denoising

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

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

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

Michael Moeller
Thomas Möllenhoff
Emanuel Laude

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

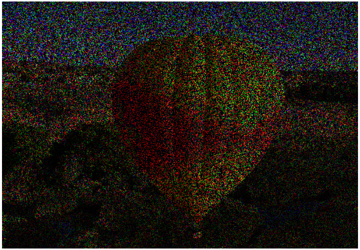
An exercise! :-)

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

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



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

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

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

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

which in this case amounts to

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

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

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

Michael Moeller
Thomas Möllenhoff
Emanuel Laude

TV-Deblurring - Option 2

The (PDHG) updates are

$$q^{k+1} = \underset{q}{\operatorname{argmin}} \frac{1}{2}\|q - (q^k + \sigma A\bar{u}^k)\|^2 + \sigma\langle f, q\rangle + \frac{\sigma}{2}\|q\|^2,$$

$$p^{k+1} = \underset{p}{\operatorname{argmin}} \frac{1}{2}\|p - (p^k + \sigma K\bar{u}^k)\|^2 + \sigma\iota_{\|\cdot\|_{2,\infty}\leq\alpha}(p),$$

$$u^{k+1} = u^k - \tau K^* p^{k+1} - \tau A^* q^{k+1}$$

$$\bar{u}^{k+1} = 2u^{k+1} - u^k.$$

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

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

PDHG implementation: Option 2 from the previous example.

Michael Moeller
Thomas Möllenhoff
Emanuel Laude

TV-Zooming

Input data

Nearest neighbor

TV Zooming

Michael Moeller
Thomas Möllenhoff
Emanuel Laude

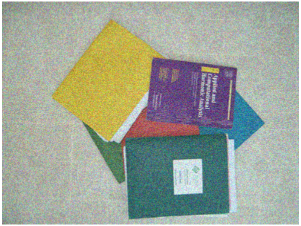
Image Segmentation

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

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

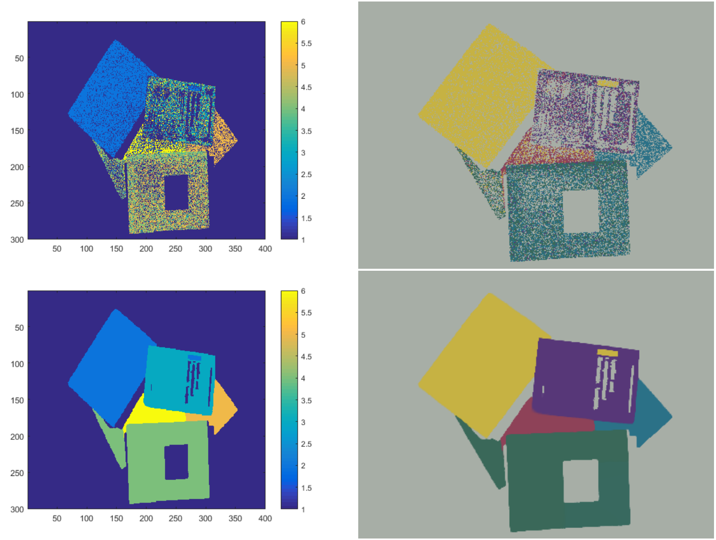
$$\iota_\Delta(u) = \begin{cases} 0 & \text{if } \sum_k u_{i,j,k} = 1, \ \forall (i,j) \\ \infty & \text{else.} \end{cases}$$

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



Michael Moeller
Thomas Möllenhoff
Emanuel Laude

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