Primal-Dual Methods

V. Estellers

WS 2017

Recall: DUALITY

Theorem (Fenchel's Duality)

Let $G : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ and $F : \mathbb{R}^m \to \mathbb{R} \cup \{\infty\}$ be proper, closed, convex functions and $u \in ri(dom(G))$ such that $Ku \in ri(dom(F))$. Then

	$\inf_u$	$G(u) + F(Ku)$	*"Primal"*
$=$	$\inf_u \sup_q$	$G(u) + \langle q, Ku \rangle - F^*(q)$	*"Saddle point"*
$=$	$\sup_q \inf_u$	$G(u) + \langle q, Ku \rangle - F^*(q)$	*"Saddle point"*
$=$	$\sup_q$	$-G^*(-K^*q) - F^*(q)$	*"Dual"*

We used the dual formulation to solve problems of the form $\min_u \ \|u - f\|_2 + \alpha \|Du\|_1$ that we could not directly because the proximal operator of $\|Du\|_1$ is not simple.

Motivation

But we still do not have a method to solve problems of the form

$$\min_u \; \|u - f\|_1 + \alpha \|Du\|_1$$

although the proximal mapping of the ℓ^1-norm is easy to compute.

Can we build an algorithm around

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

Proximal mapping as implicit gradient descent

For differentiable E, the proximal mapping does an implicit gradient step

$$u^{k+1} = \mathsf{prox}_{\tau E}(u^k) \quad \Rightarrow u^{k+1} = u^k - \tau \nabla E(u^{k+1})$$

The primal-dual hybrid gradient algorithm

Let us define

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

and try to alternate implicit ascent steps in p with implicit descent steps in u:

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

One finds

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

The primal-dual hybrid gradient algorithm

Let us define

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

and try to alternate implicit accent steps in p with implicit descent steps in u:

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

One finds

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

Primal-dual hybrid gradient method

We found

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

One should make one (currently non intuitive) modification:

Definition (PDHG)

We will call the iteration

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

the **Primal-Dual Hybrid Gradient Method**. As we will see, it converges if $\tau\sigma < \frac{1}{\|K\|^2}$.

References for PDHG

PDHG is commonly referred to as the Chambolle and Pock algorithm. Nevertheless, several authors contributed to its development.

Here is a (likely incomplete) list of relevant papers:

Pock, Cremers, Bischof, Chambolle, A convex relaxation approach for computing minimal partitions.

Esser, Zhang, Chan, A General Framework for a Class of First Order Primal-Dual Algorithms for Convex Optimization in Imaging Science.

Chambolle, Pock, A first-order primal-dual algorithm for convex problems with applications to imaging.

Zhang, Burger Osher, A unified primal-dual algorithm framework based on Bregman iteration.

Understanding PDHG

Why does PDHG work?

1. Sanity check: If the algorithm converges, it does so to a minimizer.
2. Why does PDHG converge? Computation on the board for

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

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} \in \underbrace{\begin{pmatrix} \partial G & K^T \\ -K & \partial F^* \end{pmatrix}}_{=:T} \begin{pmatrix} u^{k+1} \\ p^{k+1} \end{pmatrix} + \underbrace{\begin{pmatrix} \frac{1}{\tau} I & -K^T \\ -K & \frac{1}{\sigma} I \end{pmatrix}}_{=:M} \begin{pmatrix} u^{k+1} - u^k \\ p^{k+1} - p^k \end{pmatrix}$$

for the set-valued operator $T : \mathbb{R}^n \times \mathbb{R}^n \to \mathcal{P}(\mathbb{R}^n) \times \mathcal{P}(\mathbb{R}^n)$

Fixed point iteration

We have reformulated the update rule to

$$0 \in Tz^{k+1} + M(z^{k+1} - z^k)$$

for a set-valued operator T and a matrix M. Let us define the process of computing the next iterate as the *resolvent*

$$z^{k+1} = (M + T)^{-1}(Mz^k). \qquad \text{(CPPA)}$$

We already know an iteration of this form, the proximal point algorithm

$$u^{k+1} = prox_E(u^k) = (I + \tau\partial E)^{-1}(u^k)$$

So we can use the same tools to analyze its convergence. We will call it a *customized proximal point algorithm* (CPPA).

Convergence of the CPPA

Remember what we did for the proximal gradient algorithm?

$\rightarrow$ Show that $prox_E = (I + \tau \partial E)^{-1}$ is firmly nonexpansive, i.e. averaged with $\alpha = 1/2$.

We will do something similar by generalizing the crucial inequality

$$\langle p_u - p_v, u - v \rangle \geq 0 \qquad \forall u, v, p_u \in \partial E(u), p_v \in \partial E(v)$$

Definition (Monotone Operator)

A set valued operator T is called *monotone* if the inequality

$$\langle p_u - p_v, u - v \rangle \geq 0$$

holds for all u, v, $p_u \in T(u)$ and $p_v \in T(v)$.

Convergence of the CPPA

This has the potential to show convergence of

$$0 \in T(z^{k+1}) + z^{k+1} - z^k, \qquad \text{(PPA)}$$

provided that the above iteration is well-defined, i.e. the resolvent $(I+T)^{-1}(z)$ is defined for any $z \in \mathbb{R}^n$. This is a technical issue which can be resolved by considering *maximal* monotone operators. In our convex settings, this is not an issue.

Definition

The relation T is **maximal monotone** if there is no monotone operator that properly contains it as a subset of $\mathbb{R}^n \times \mathbb{R}^n$.

In other words, if the monotone operator T is not maximal, then there is $(x, u) \notin T$ such that $T \cup \{(x, u)\}$ is monotone.

Examples of maximal monotone operators

Lemma

$E\colon \mathbb{R}^n \to \bar{\mathbb{R}}$, then ∂E is a monotone operator. If E is closed convex and proper then ∂E is maximal monotone.

Lemma

A continuous monotone function $F\colon \mathbb{R}^n \to \mathbb{R}$ with dom$(F) = \mathbb{R}^n$ is maximal.

Lemma

If T is maximal monotone, then the resolvent $R_T = (I + \alpha T)^{-1}$ with $\alpha > 0$ and the Caley operator $C_T = 2R_T - I$ are nonexpansive functions.

Theorem (Convergence of Generalized Proximal Point Algorithm)

Let T be a maximal monotone operator, and let there exist a z such that $0 \in T(z)$. Then the (generalized) proximal point algorithm

$$\begin{aligned} z^{k+1} &= (T+I)^{-1}(z^k) \\ 0 &\in T(z^{k+1}) + z^{k+1} - z^k \end{aligned} \tag{2}$$

converges to a point $\tilde{z}$ with $0 \in T(\tilde{z})$.

Proof.

If T is maximal monotone, the resolvent $R_T = (T+I)^{-1}$ and the Caley operator $C_T = 2R_T - I$ are nonexpansize. Since $R_T = \frac{1}{2}I + \frac{1}{2}C_T$, the resolvent R_T is an averaged operator and the generalized proximal point algorithm is a fixed-point iteration of an averaged operator that converges by Krasnoselskii-Mann Theorem. □

Convergence of the CPPA

But we wrote the PDHG algorithm as

$$0 \in T(z^{k+1}) + Mz^{k+1} - Mz^k, \tag{3}$$

i.e. with an additional matrix M.
Idea: For symmetric positive definite matrices, write $M = L^T L$ and rewrite (CPPA) as

$$0 \in L^{-T} T L^{-1}(\zeta^{k+1}) + \zeta^{k+1} - \zeta^k, \tag{CPPA}$$

with $\zeta^k = Lz^k$, and

$$L^{-T} T L^{-1}(\zeta) = \{q \in \mathbb{R}^n \mid q = L^{-T} p, \;\; p \in T(L^{-1}\zeta)\}.$$

Lemma

If T is monotone, then $L^{-T} T L^{-1}$ is monotone, too.

Proof: Exercise.

Convergence conclusions CPPA

Theorem (Convergence CPPA)

Let T be a maximally monotone operator. Let there exist a z such that $0 \in T(z)$, and let the matrix M be symmetric positive definite. Then the customized proximal point algorithm

$$z^{k+1} = (M + T)^{-1}(Mz^k)$$

converges to a $\hat{z}$ with $0 \in T(z)$.

Convergence conclusions PDHG

As the primal-dual hybrid gradient method can be rewritten (after an index shift) as

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} \in \underbrace{\begin{pmatrix} \partial G & K^T \\ -K & \partial F^* \end{pmatrix}}_{=:T} \begin{pmatrix} u^{k+1} \\ p^{k+1} \end{pmatrix} + \underbrace{\begin{pmatrix} \frac{1}{\tau} I & -K^T \\ -K & \frac{1}{\sigma} I \end{pmatrix}}_{=:M} \begin{pmatrix} u^{k+1} - u^k \\ p^{k+1} - p^k \end{pmatrix}.$$

Theorem

Convergence PDHG The operator T is maximally monotone. For $\tau\sigma < \frac{1}{\|K\|^2}$ the matrix M in the PDHG algorithm is positive definite. Hence, PDHG converges.

(Assuming F and G to be proper, closed, and convex, assuming there is a $u \in$ ri(G) such that $Ku \in$ ri(F), and assuming the existence of a minimizer).

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

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

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

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

which in this case amounts to

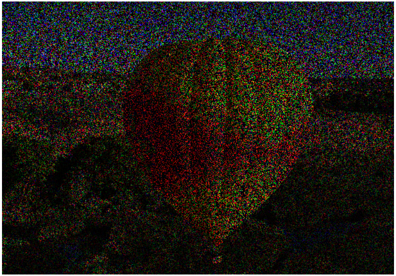
An exercise! :-)

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

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

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.

TV-Deblurring - Option 1

We write

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

The (PDHG) updates are

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

which in this case amounts to

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

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

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

TV-Deblurring - Option 2

The (PDHG) updates are

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

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.

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

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

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

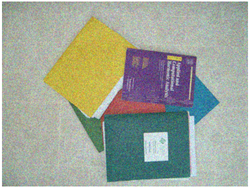
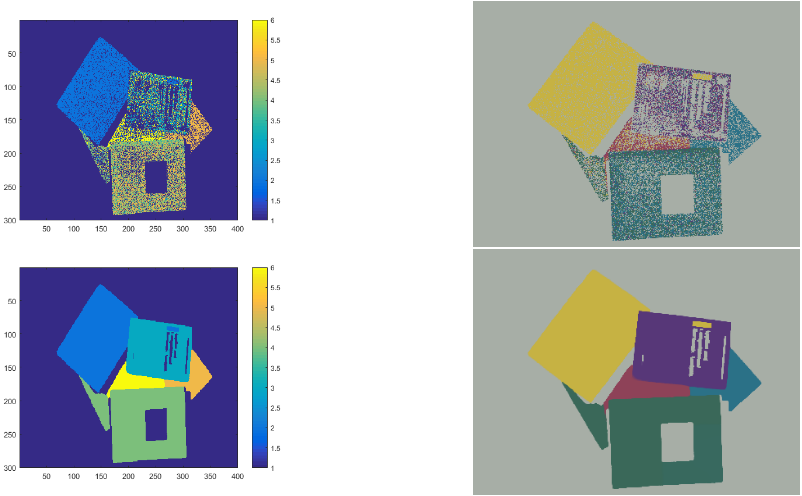


Image Segmentation



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

Image Segmentation

Option 1: We solve

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

→ 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 $(Su)_{i,j} = \sum_k u_{i,j}$.

→ Very simple proximal operators, but additional variable.

Final remark for applications

If you are too lazy to compute the proximity operator of F^*

$$\begin{aligned}\tilde{p} &= \mathrm{prox}_{\sigma F^*}(z)\\ &= \arg\min_p \frac{1}{2}\|p-z\|^2 + \sigma F^*(p)\\ \Rightarrow 0 &= \tilde{p} - z + \sigma \tilde{u}, \quad \tilde{u} \in \partial F^*(\tilde{p})\\ \Rightarrow 0 &= \tilde{u} - z/\sigma + \frac{1}{\sigma}\tilde{p}, \quad \tilde{p} \in \partial F(\tilde{u})\\ \Rightarrow \tilde{u} &= \mathrm{prox}_{\frac{1}{\sigma}F}(z/\sigma)\\ \Rightarrow \tilde{p} &= z - \sigma\ \mathrm{prox}_{\frac{1}{\sigma}F}(z/\sigma)\end{aligned}$$

Lemma (Moreau's identity)

If you know prox_F *you also know* prox_{F^*},

$$\mathrm{prox}_{\sigma F^*}(z) = z - \sigma\ \mathrm{prox}_{\frac{1}{\sigma}F}(z/\sigma).$$

Convergence rate

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 a convergence rate of

$$\min_{j\in\{0,\ldots,k\}} \|(I + L^{-T}TL^{-1})(\xi^k) - \xi^k\|^2 \leq C\frac{\|\xi^0 - \xi^0\|}{k+1}$$

for $\xi^k = L(u^k, p^k)$, L being the matrix square-root of M, and C being a constant.

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

Either G or F^* is strongly convex

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

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

Theorem

For strongly convex G and $\epsilon > 0$, there exists an N_0 such that for any $N \geq N_0$:

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

Discussion of the convergence orders

If part of the energy is L smooth, the gradient methods obtain linear convergence on strongly convex energies.

As L-smoothness of the primal corresponds to $1/L$-strong convexity of the convex conjugate. It is natural to ask: what can we do if we additionally assume F to be L-smooth, i.e., assume F^* to be strongly convex?

Two strongly convex functions

Consider

$$\begin{aligned} p^{k+1} &= \mathsf{prox}_{\sigma F^*}(p^k + \sigma K \bar{u}^k), \\ u^{k+1} &= \mathsf{prox}_{\tau G}(u^k - \tau K^* p^{k+1}), \\ \bar{u}^{k+1} &= u^{k+1} + \theta(u^{k+1} - u^k). \end{aligned} \tag{PDHG3}$$

Theorem (Linear convergence of strongly convex functions)

For $\mu \le 2\sqrt{\gamma\delta}/\|K\|$, $\tau = \mu/(2\gamma)$, $\sigma = \mu/(2\delta)$, $\theta \in [1/(1+\mu), 1]$, G *being* γ*-strongly convex and* F^* *being* δ*-strongly convex, there exists* $\omega < 1$, *such that the iterates of* (PDHG3) *meet*

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

Summary: descent methods

For energies of the form

$$u^* \in \arg\min_{u \in \mathbb{R}^n} F(u) + G(u),$$

for proper, closed, convex $F : \mathbb{R}^n \to \mathbb{R}$, $G : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$, with F additionally being L-smooth, we discussed

Gradient descent: $G \equiv 0$

Gradient projection: $G = \delta_C$

Proximal gradient: G simple (easy to compute prox)

Convex Conjugate: Geometric interpretation

Definition

Let $E : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ be any function, not necessarily convex, we define its *convex conjugate* to be

$$E^*(p) = \sup_{u \in \mathbb{R}^n} [\langle u, p \rangle - E(u)] .$$

For $E\colon \mathbb{R} \mapsto \mathbb{R}$ and $p \in \mathbb{R}$, the conjugate function $E^*(p)$ is the maximum gap between the linear function pu and $E(u)$ (dashed line).

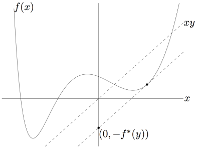


Fig.: Source: Boyd, and Vandenberghe. *Convex optimization theory*.2004

Convex Conjugate: Geometric interpretation

Definition

Let $E : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ be any function, not necessarily convex, we define its *convex conjugate* to be

$$E^*(p) = \sup_{u \in \mathbb{R}^n} \left[\langle u, p \rangle - E(u) \right].$$

The points of the epigraph of E^* parameterize the affine functions minorizing E.

$$\begin{aligned} (p, \alpha) \in \mathsf{epi}(E^*) &\iff \alpha \geq \sup_{u \in \mathbb{R}^n} \left[\langle u, p \rangle - E(u) \right] \\ &\iff E(u) \geq \langle u, p \rangle - \alpha \quad \forall u \in \mathbb{R}^n. \end{aligned} \tag{4}$$

If the affine function $l(u) = \langle p, u - \alpha$ minorizes E, then the affine function $m(u) = \langle p, u - E^*(p)$ is the largest affine minorizer and satisfies

$$l(u) \leq m(u) \leq E(u).$$

Biconjugate

Lemma

The convex conjugate of any proper function $E : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ is closed (or lower semi-continuous) and convex.

Note that $E^{**}(u) \leq E(u)$ because

$$\begin{aligned} E^{**}(u) &= \sup_p \langle p, u \rangle - E^*(p) = \sup_p \langle p, u \rangle - \sup_v [\langle p, v \rangle - E(v)] \\ &\leq \sup_p \langle p, u \rangle - [\langle p, u \rangle - E(u)] = E(u). \end{aligned}$$

E^{**} is the largest convex lower semi-continuous envelope of E.

- conjugation reverses inequalities: if $E \leq F$ then $E^* \geq F^*$.
- The conjugate function is always convex and lower semi-continuous.
- If E is proper, convex, and lower semi-continuous, then $E^{**} = E$.

If $\hat{E}$ is a convex lower semi-continuous function s.t. $\hat{E} \leq E$, then $\hat{E}^{**} = \hat{E} \leq E^{**}$.

Duality

We showed that for E proper, convex and closed, $E^{**} = E$.
Let $G : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ and $F : \mathbb{R}^m \to \mathbb{R} \cup \{\infty\}$ be proper, closed, convex functions. Consider $F(u) = F^{**}(u) = \sup_p \langle p, u \rangle - F^*(p)$ in

$$\inf_u \; G(u) + F(Ku)$$

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

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

Switch inf and sup and apply $G^*(-K^*p) = \sup_u \; -[\langle K^*p, u \rangle + G(u)]$

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

$$\sup_p -F^*(p) + \inf_u \; G(u) + \langle K^*p, u \rangle$$

$$\sup_p -F^*(p) - \underbrace{\sup_u \; -[G(u) + \langle K^*p, u \rangle]}_{G^*(K^*p)}$$

Duality

Theorem (Fenchel's Duality[1])

Let $G : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ and $F : \mathbb{R}^m \to \mathbb{R} \cup \{\infty\}$ be proper, closed, convex functions and let there exist a $u \in$ ri$($dom$(G))$ such that $Ku \in$ ri$($dom$(F))$. Then

$$
\begin{array}{llll}
 & \inf_u & G(u) + F(Ku) & \textit{"Primal"} \\
= & \inf_u \sup_q & G(u) + \langle q, Ku \rangle - F^*(q) & \textit{"Saddle point"} \\
= & \sup_q \inf_u & G(u) + \langle q, Ku \rangle - F^*(q) & \textit{"Saddle point"} \\
= & \sup_q & -G^*(-K^*q) - F^*(q) & \textit{"Dual"}
\end{array}
$$

[1] C.f. Rockafellar, *Convex Analysis*, Section 31

Motivation for dual algorithms

With the descent algorithms from the previous lecture, we could not solve

$$u^* \in \operatorname*{argmin}_u \frac{1}{2}\|u - f\|_2^2 + \alpha\|Du\|_{2,1}$$

b we can solve its dual

$$\hat{q} = \operatorname*{argmax}_{\|q\|_{2,\infty}\leq 1} -\frac{1}{2}\|\alpha D^* q - f\|_2^2 = \operatorname*{argmin}_{\|q\|_{2,\infty}\leq 1} \frac{1}{2}\left\|D^* q - \frac{f}{\alpha}\right\|_2^2$$

$$\hat{q} = \operatorname*{argmin}_{q\in C} \frac{1}{2}\left\|D^* q - \frac{f}{\alpha}\right\|_2^2$$

where $C = \{q \in \mathbb{R}^{nm\times 2c} \mid \|q\|_{2,\infty} \leq 1\}$ with projected gradient descent.
Recover $\hat{u}$ from optimality conditions of saddle-point $-K^T\hat{p} \in \partial G(\hat{u})$.

Motivation for primal-dual algorithms

With the descent algorithms from the previous lecture, we could not solve
Consider TV-ℓ^1 denoising, i.e.,

$$\inf_u \ \|u - f\|_1 + \alpha \|Du\|_{2,1}$$

nor its dual

$$\sup_q \ \langle \alpha D^* q, f \rangle - \delta_{\|\cdot\|_\infty \leq 1}(-\alpha D^* q) - \delta_{\|\cdot\|_{2,\infty} \leq 1}(q)$$

and directly worked with the saddle-point formulation

$$\inf_u \sup_q \|u - f\|_1 + \alpha \langle q, Du \rangle - \delta_{\|\cdot\|_{2,\infty} \leq 1}(q)$$

$$\inf_u \sup_q G(u) + \langle q, Ku \rangle - F^*(q)$$

Generic form Primal-Dual Algorithm

Remember the optimality conditions of the saddle point formulation

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

were

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} \in \underbrace{\begin{pmatrix} \partial G & K^T \\ -K & \partial F^* \end{pmatrix}}_{T} \underbrace{\begin{pmatrix} \hat{u} \\ \hat{p} \end{pmatrix}}_{\hat{z}}$$

$$0 \in T(\hat{z})$$

We could not compute $\hat{z}$ directly. Therefore, given $M \succ 0$ we defined the iteration $T(z^{k+1}) + M(z^{k+1} - z^k)$ such that

$$\hat{z} \;\text{ is a fixed point of }\; 0 \in T(z^{k+1}) + M(z^{k+1} - z^k) \iff 0 \in T(\hat{z})$$

Generic form Primal-Dual Algorithm

In terms of $(\hat{u}, \hat{p}) = \hat{z}$ we have

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} \in \begin{pmatrix} \partial G & K^T \\ -K & \partial F^* \end{pmatrix} \begin{pmatrix} u^{k+1} \\ p^{K+1} \end{pmatrix} \quad + \overbrace{\begin{pmatrix} M_1 & M_3 \\ M_4 & M_2 \end{pmatrix}}^{=:M} \begin{pmatrix} u^{k+1} - u^k \\ p^{K+1} - p^k \end{pmatrix}$$

$$\begin{aligned} 0 &\in \partial G(u^{k+1}) + K^T p^{k+1} && + M_1(u^{k+1} - u^k) + M_3(p^{k+1} - p^k) \\ 0 &\in -K u^{k+1} + \partial F^*(p^{k+1}) && + M_4(u^{k+1} - u^k) + M_2(p^{k+1} - p^k) \end{aligned}$$

for sequential updates, we set $M_3 = -K^T$, or $M_4 = K$

for a symmetric M, we set $M_3 = (M_4)^T$,

for $M \succ 0$, many options. In PDHG $M_1 = \frac{1}{\tau} I$, $M_2 = \frac{1}{\sigma} I$

The choice of M_1 and M_2 does not influence the convergence rate

ADMM

Let us consider

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} \in \begin{pmatrix} \partial G & K^T \\ -K & \partial F^* \end{pmatrix} \begin{pmatrix} u^{k+1} \\ p^{K+1} \end{pmatrix} + \underbrace{\begin{pmatrix} \frac{1}{\lambda} I & -K^T \\ -K & \lambda K K^T \end{pmatrix}}_{=: M} \begin{pmatrix} u^{k+1} - u^k \\ p^{K+1} - p^k \end{pmatrix}.$$

The resulting M is only positive semi-definite. Exploit fixed point iterations of averaged operators in a different way to show convergence.

If we decompose this equation component by component, in u we have

$$\begin{aligned} 0 &\in \partial G(u^{k+1}) + \frac{1}{\lambda}(u^{k+1} - u^k) + K^T p^k \\ 0 &\in \lambda \partial G(u^{k+1}) + u^{k+1} - (u^k - \lambda K^T p^k) \\ u^{k+1} &= \operatorname*{argmin}_u \ \lambda G(u) + \frac{1}{2}\|u - (u^k - \lambda K^T p^k)\|^2 \\ u^{k+1} &= \mathsf{prox}_{\lambda G}(u^k - \lambda K^T p^k), \end{aligned}$$

which requires a proximal step to update the primal variable, like PDHG.

ADMM

Let us consider

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} \in \begin{pmatrix} \partial G & K^T \\ -K & \partial F^* \end{pmatrix} \begin{pmatrix} u^{k+1} \\ p^{K+1} \end{pmatrix} + \underbrace{\begin{pmatrix} \frac{1}{\lambda} I & -K^T \\ -K & \lambda K K^T \end{pmatrix}}_{=:M} \begin{pmatrix} u^{k+1} - u^k \\ p^{K+1} - p^k \end{pmatrix}.$$

If we decompose this equation component by component, in p we have

$$0 \in \partial F^*(p^{k+1}) + \lambda K K^T (p^{k+1} - p^k) - K(2u^{k+1} - u^k)$$

$$p^{k+1} = \underset{p}{\operatorname{argmin}} \; F^*(p) + \frac{\lambda}{2} \left\| K^T p - K^T p^k - \frac{1}{\lambda} K(2u^{k+1} - u^k) \right\|^2,$$

and we need a special structure of K or F^* to solve this subproblem because. In general, the subproblem is more difficult than the proximal step of PDHG.

Example: Graph projection splitting

For any generic problem of our usual form,

$$\min_u H(u) + R(Du)$$

we can write

$$\min_{u,v,d} \; H(v) + R(d), \qquad \text{s.t.} \underbrace{\begin{pmatrix} I & -I & 0 \\ D & 0 & -I \end{pmatrix}}_{K} \underbrace{\begin{pmatrix} u \\ v \\ d \end{pmatrix}}_{x} = 0$$

$$\min_{x=(u,v,d)} \; H(v) + R(d) + F(Kx) \quad F(z) = \begin{cases} 0 & \text{if } z = 0 \\ \infty & \text{otherwise} \end{cases}$$

$$\min_{x=(u,v,d)} \max_p \; H(v) + R(d) + \langle Kx, p \rangle$$

where $F^* = \sup_z \langle z, p \rangle - F(z) = \langle 0, p \rangle - F(0) = 0$ and the solution of the ADMM subproblem in p becomes a linear system.

Remarks on ADMM

ADMM is often derived from a different perspective. In this perspective, the above ADMM is the classical algorithm applied to the dual formulation of the problem. The primal version is

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} \in \begin{pmatrix} \partial G & K^T \\ -K & \partial F^* \end{pmatrix} \begin{pmatrix} u^{k+1} \\ p^{K+1} \end{pmatrix} + \underbrace{\begin{pmatrix} \lambda K^T K & K^T \\ K & \frac{1}{\lambda} I \end{pmatrix}}_{=:M} \begin{pmatrix} u^{k+1} - u^k \\ p^{K+1} - p^k \end{pmatrix}.$$

and requires G to be sufficiently simple in order to solve the update equations, i.e.

$$\begin{aligned} p^{k+1} &= \mathsf{prox}_{\lambda F^*}(p^k + \lambda K u^k) \\ u^{k+1} &= \arg\min_u \ \frac{\lambda}{2} \left\| Ku - Ku^k + \frac{1}{\lambda}(2p^{k+1} - p^k) \right\|^2 + G(u) \end{aligned}$$

Some final remarks

Detailed convergence rate of ADMM is still an active field of research.

Whether or not ADMM is faster than PDHG and its variants largely depends on how efficient the non-prox step can be computed.
It often even depends on the architecture you are computing on.
Tendency:

PDHG is better parallelizable $\to$ GPU

ADMM makes more progress per iteration $\to$ CPU

Stopping customized proximal point algorithms

We find a point $(\hat{u}, \hat{p})$ that satisfies the optimality conditions

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} \in \begin{pmatrix} \partial G & K^T \\ -K & \partial F^* \end{pmatrix} \begin{pmatrix} \hat{u} \\ \hat{p} \end{pmatrix}$$

of the saddle point problem

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

by the customized proximal point algorithm

$$0 \in \begin{bmatrix} \partial G & K^T \\ -K & \partial F^* \end{bmatrix} \begin{bmatrix} u^{k+1} \\ p^{k+1} \end{bmatrix} + \overbrace{\begin{bmatrix} M_1 & -K^T \\ -K & M_2 \end{bmatrix}}^{M \succ 0} \begin{bmatrix} u^{k+1} - u^k \\ p^{k+1} - p^k \end{bmatrix}$$

Natural considerations:

How close is $-K^T p^{k+1}$ to being an element of $\partial G(u^{k+1})$?

How close is Ku^{k+1} to being an element of $\partial F^*(p^{k+1})$?

How close is (u^{k+1}, p^{k+1}) to (u^k, p^k) with distance induced by M?

Primal and dual residuals

Based on the **primal** *and* **dual** *residuals*:

$$r_p^{k+1} = M_2(p^{k+1} - p^k) - K(u^{k+1} - u^k)$$
$$r_d^{k+1} = M_1(u^{k+1} - u^k) - K^T(p^{k+1} - p^k)$$

we consider our algorithm to be convergent if $\|r_d^{k+1}\|^2 + \|r_p^{k+1}\|^2 \to 0$, because this implies

$$\text{dist}(-K^T p^{k+1}, \partial G(u^{k+1})) \to 0, \quad \text{dist}(Ku^{k+1}, \partial F^*(p^{k+1})) \to 0.$$

This does not imply convergence of u^k and p^k by itself, but as we know that PDHG and ADMM do converge, then $\|r_d^{k+1}\|$ and $\|r_p^{k+1}\|$ are good measures for convergence.

Upper bounds on the residuals

How should we use $\|r_d^{k+1}\|$ and $\|r_p^{k+1}\|$ to formalize a stopping criterion?

Simple option: Iterator until $\|r_d^{k+1}\| \leq \epsilon$ and $\|r_p^{k+1}\| \leq \epsilon$.

Could be unfair, if $u^k \in \mathbb{R}^n$ and $p^k \in \mathbb{R}^m$ and e.g. $n >> m$.
Use $\|r_d^{k+1}\| \leq \sqrt{n}\ \epsilon$ and $\|r_p^{k+1}\| \leq \sqrt{m}\ \epsilon$.

Could be unfair for different scales! Introduce absolute and relative error criteria:

$$\begin{aligned} \|r_d^{k+1}\| &\leq \sqrt{n}\ \epsilon^{abs} + \text{dual scale factor} \cdot \epsilon^{rel} \\ \|r_p^{k+1}\| &\leq \sqrt{m}\ \epsilon^{abs} + \text{primal scale factor} \cdot \epsilon^{rel} \end{aligned}$$

But what are reasonable scale factors?

Scaling the primal residuum

$$0 \in \begin{bmatrix} \partial G & K^T \\ -K & \partial F^* \end{bmatrix} \begin{bmatrix} u^{k+1} \\ p^{k+1} \end{bmatrix} + \begin{bmatrix} M_1 & -K^T \\ -K & M_2 \end{bmatrix} \begin{bmatrix} u^{k+1} - u^k \\ p^{k+1} - p^k \end{bmatrix}$$

The primal residual

$$r_p^{k+1} = M_2(p^{k+1} - p^k) - K(u^{k+1} - u^k)$$

measures how far Ku^{k+1} is away from a particular element in $\partial F^*(p^{k+1})$, and therefore scales with the magnitude of elements in $\partial F^*(p^{k+1})$.

$$\begin{aligned} & 0 \in \partial F^*(p^{k+1}) - Ku^{k+1} + r_p^{k+1} \\ \Rightarrow & 0 \in \partial F^*(p^{k+1}) - K^T(2u^{k+1} - u^k) + M_2(p^{k+1} - p^k). \\ \Rightarrow & \underbrace{M_2(p^k - p^{k+1}) + K^T(2u^{k+1} - u^k)}_{=:z^{k+1}} \in \partial F^*(p^{k+1}) \end{aligned}$$

Thus, $\|r_p^{k+1}\| \leq \sqrt{m}\ \epsilon^{abs} + \|z^{k+1}\| \cdot \epsilon^{rel}$ is scale-independent.

Scaling the dual residuum

Similarly, the dual residual

$$r_d^{k+1} = M_1(u^{k+1} - u^k) - K^T(p^{k+1} - p^k)$$

measures how far $-K^T p^{k+1}$ is away from a particular element in $\partial G(u^{k+1})$, and scales with the magnitude of elements in $\partial G(u^{k+1})$.

$$\begin{aligned} & 0 \in \partial G(u^{k+1}) + K^T p^{k+1} + r_d^{k+1}. \\ \Rightarrow\ & 0 \in \partial G(u^{k+1}) + K^T p^k + M_1(u^{k+1} - u^k) \\ \Rightarrow\ & \underbrace{M_1(u^k - u^{k+1}) - K^T p^k}_{=:v^{k+1}} \in \partial G(u^{k+1}) \end{aligned}$$

Thus, $\|r_d^{k+1}\| \leq \sqrt{n}\ \epsilon^{abs} + \|v^{k+1}\| \cdot \epsilon^{rel}$ is scale-independent.

A scaled absolute and relative stopping criterion

In summary, a good stopping criterion is

$$\|r_p^{k+1}\| \leq \sqrt{m}\ \epsilon^{abs} + \|z^{k+1}\| \cdot \epsilon^{rel},$$
$$\|r_d^{k+1}\| \leq \sqrt{n}\ \epsilon^{abs} + \|v^{k+1}\| \cdot \epsilon^{rel}.$$

Interesting observation in our previous considerations:
ADMM/PDHG actually generates iterates $(u^{k+1}, p^{k+1}, v^{k+1}, z^{k+1})$ with

$$v^{k+1} \in \partial G(u^{k+1}), \qquad z^{k+1} \in \partial F^*(p^{k+1}).$$

The goal of all algorithms is to achieve convergence

$$\| \underbrace{z^{k+1} - Ku^{k+1}}_{=r_p^{k+1}} \| \rightarrow 0 \text{ and } \| \underbrace{v^{k+1} + K^T p^{k+1}}_{=r_d^{k+1}} \| \rightarrow 0!$$

Adaptive stepsizes

r_p^{k+1} and r_d^{k+1} determine the convergence of the algorithm.

Can we also use r_d and r_p to accelerate the algorithm?

Adaptive stepsizes:

$$0 \in \begin{bmatrix} \partial G & K^T \\ -K & \partial F^* \end{bmatrix} \begin{bmatrix} u^{k+1} \\ p^{k+1} \end{bmatrix} + \begin{bmatrix} \frac{1}{\tau^k} M_1 & -K^T \\ -K & \frac{1}{\sigma^k} M_2 \end{bmatrix} \begin{bmatrix} u^{k+1} - u^k \\ p^{k+1} - p^k \end{bmatrix}$$

Base the choices of τ^k and σ^k on the residuals r_p^k and r_d^k, where

$$\begin{aligned} r_p^{k+1} &= \frac{1}{\sigma^k} M_2(p^{k+1} - p^k) - K(u^{k+1} - u^k), \\ r_d^{k+1} &= \frac{1}{\tau^k} M_1(u^{k+1} - u^k) - K^T(p^{k+1} - p^k)? \end{aligned}$$

Customized proximal point algorithms

Decreasing residual balancing: Let $(M_1, -K^T; -K, M_2)$ be positive definite. Pick τ^0 and σ^0 with $\tau^0\sigma^0 < 1$. Further choose $\mu > 1$, $\alpha^0 < 1$, $\beta < 1$ and adapt as follows

If $\|r_p^k\| > \mu\|r_d^k\|$, do

$$\tau^{k+1} = (1-\alpha^k)\tau^k, \quad \sigma^{k+1} = \frac{1}{1-\alpha^k}\sigma^k, \quad \alpha^{k+1} = \alpha^k \cdot \beta.$$

If $\|r_d^k\| > \mu\|r_p^k\|$, do

$$\tau^{k+1} = \frac{1}{1-\alpha^k}\tau^k, \quad \sigma^{k+1} = (1-\alpha^k)\sigma^k, \quad \alpha^{k+1} = \alpha^k \cdot \beta.$$

Keep $\tau^{k+1} = \tau^k$, $\sigma^{k+1} = \sigma^k$, and $\alpha^{k+1} = \alpha^k$ otherwise.

Goldstein et al., *Adaptive Primal-Dual Hybrid Gradient Methods for Saddle-Point Problems*: The resulting scheme still converges.

Summary

For proper, closed, convex functions G and $F \circ K$ (with $\mathrm{ri}(\mathrm{dom}(G)) \cap \mathrm{ri}(\mathrm{dom}(F \circ K)) \neq \emptyset$) we can write

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

with the optimality condition

$$0 \in \begin{bmatrix} \partial G & K^T \\ -K & \partial F^* \end{bmatrix} \begin{bmatrix} \hat{u} \\ \hat{p} \end{bmatrix}.$$

Typically, $(\hat{u}, \hat{p})$ cannot be computed directly, but iterative methods on this saddle point problem that decouple the update inclusions in u and p.

They converge analyzed as fixed-point iterations of averaged operators.

Saddle point methods

Most prominently, we discussed

PDHG, overrelaxation on primal

$$0 \in \begin{bmatrix} \partial G & K^T \\ -K & \partial F^* \end{bmatrix} \begin{bmatrix} u^{k+1} \\ p^{k+1} \end{bmatrix} + \begin{bmatrix} \frac{1}{\tau} I & -K^T \\ -K & \frac{1}{\sigma} I \end{bmatrix} \begin{bmatrix} u^{k+1} - u^k \\ p^{k+1} - p^k \end{bmatrix}.$$

PDHG, overrelaxation on dual

$$0 \in \begin{bmatrix} \partial G & K^T \\ -K & \partial F^* \end{bmatrix} \begin{bmatrix} u^{k+1} \\ p^{k+1} \end{bmatrix} + \begin{bmatrix} \frac{1}{\tau} I & K^T \\ K & \frac{1}{\sigma} I \end{bmatrix} \begin{bmatrix} u^{k+1} - u^k \\ p^{k+1} - p^k \end{bmatrix}.$$

Primal ADMM

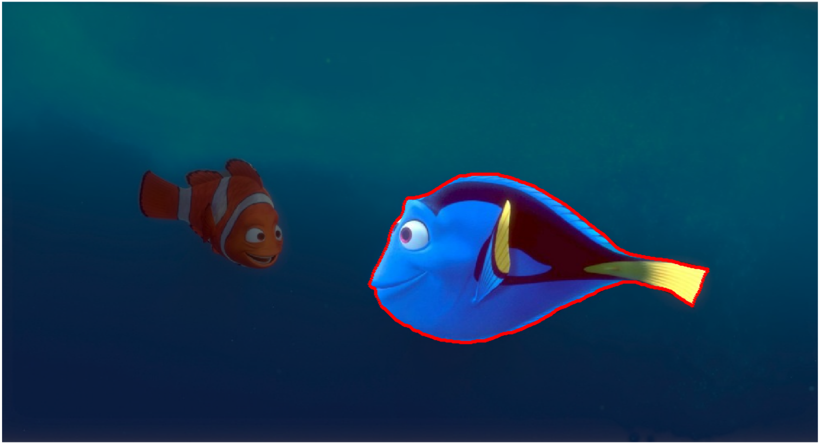
$$0 \in \begin{bmatrix} \partial G & K^T \\ -K & \partial F^* \end{bmatrix} \begin{bmatrix} u^{k+1} \\ p^{k+1} \end{bmatrix} + \begin{bmatrix} \lambda K^T K & K^T \\ K & \frac{1}{\lambda} I \end{bmatrix} \begin{bmatrix} u^{k+1} - u^k \\ p^{k+1} - p^k \end{bmatrix}.$$

Corresponding dual ADMM

$$0 \in \begin{bmatrix} \partial G & K^T \\ -K & \partial F^* \end{bmatrix} \begin{bmatrix} u^{k+1} \\ p^{k+1} \end{bmatrix} + \begin{bmatrix} \frac{1}{\lambda} I & -K^T \\ -K & \lambda K K^T \end{bmatrix} \begin{bmatrix} u^{k+1} - u^k \\ p^{k+1} - p^k \end{bmatrix}.$$

Single view 3D reconstruction

Let Ω be the image domain, $S \subset \Omega$ an object.



From: Finding Nemo, https://ohmy.disney.com/movies/2015/12/20/dory-finding-nemo-hero/

Goal: Estimate a 3D model

First version: Single view 2.5D reconstruction

Oswald, Töppe, Cremers CVPR 2012: Find a height map that has minimal surface for fixed volume and respects the contour.

Mathematically for height map $u : S \to \mathbb{R}$

- $\int_S u(x)\, dx = V$, where V is a user given volume
- Constrain $u_{|\partial S} = 0$
- Minimize $\int_S \sqrt{1 + |\nabla u(x)|^2}\, dx$ (surface area)

Discrete form

$$\min_u \quad \sum_i \sqrt{1 + |(Du)_i|^2} + \delta_{\Sigma_V}(u),$$

for a suitable gradient operator D (respecting $u_{|\partial S} = 0$),

$$\Sigma_V = \{u \in \mathbb{R}^{|S|} \mid \sum_i u_i = V\}.$$

Single view 2.5D reconstruction

How can we minimize

$$E(u) = \sum_i \sqrt{1 + |(Du)_i|^2} + \delta_{\Sigma_V}(u) \text{ ?}$$

One option: Gradient projection.

Descent on the term that does not have an easy prox:

$$u^{k+1/2} = u^k - \tau D^* v^k, \qquad v_{i,:} = \frac{(Du^k)_{i,:}}{\sqrt{1 + |(Du^k)_{i,:}|^2}}$$

for suitable τ, with $D : \mathbb{R}^n \to \mathbb{R}^{n \times 2}$.

Project onto constraint set:

$$\mathsf{proj}_{\Sigma_V}(v) = \operatorname*{argmin}_u \frac{1}{2}\|u - v\|_2^2 + \delta_{\Sigma_V}(u)$$

How does the projection look like?

Single view 2.5D reconstruction

$$\operatorname*{argmin}_u \frac{1}{2}\|u-v\|_2^2 + \delta_{\Sigma_V}(u) = \operatorname*{argmin}_u \frac{1}{2}\|u-v\|_2^2 + \delta_{\cdot - V}(\langle \mathbf{1}, u\rangle)$$

Optimality condition

$$0 = \hat{u} - v + \mathbf{1}p, \qquad p \in \partial\delta_{\cdot - V}(\langle \mathbf{1}, \hat{u}\rangle)$$

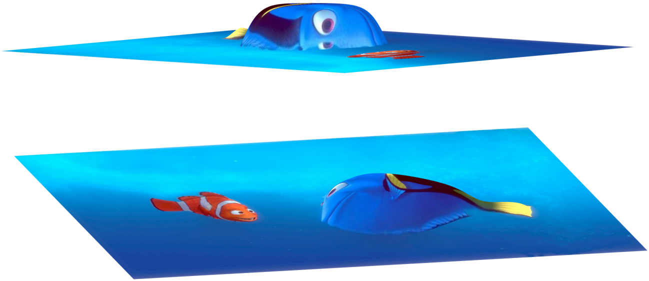
Taking the inner product with $\mathbf{1}$ and recalling that $\sum_i \hat{u}_i = V$:

$$0 = V - \sum_i v_i + np \Rightarrow p \qquad = \frac{1}{n}\left(V - \sum_i v_i\right),$$

yields

$$\hat{u} = v - \mathbf{1}\frac{1}{n}\left(V - \sum_i v_i\right) = v - \mathsf{mean}(v)\mathbf{1} + \mathbf{1}\frac{V}{n}$$

Single view 2.5D reconstruction



Oringinal image from: Finding Nemo, `https://ohmy.disney.com/movies/2015/12/20/dory-finding-nemo-hero/`

Single view 2.5D reconstruction

What about our primal-dual/splitting methods?

$$\min_{u} \quad \sum_i \sqrt{1 + |(Du)_i|^2} + \delta_{\Sigma_V}(u),$$

Natural reformulation:

$$\min_{u,d} \quad \sum_i \sqrt{1 + |d_i|^2} + \delta_{\Sigma_V}(u), \quad Du = d.$$

But is $F(d) = \sum_i \sqrt{1 + |d_i|^2}$ simple?

Somewhat yes, as it reduces to a 1D problem.

Somewhat no, as there is no (easy) closed form solution.

Reformulation that makes the prox operator really easy?

Single view 2.5D reconstruction

Let's start with

$$\min_{u,d} \quad \sum_i \sqrt{1 + |d_i|^2} + \delta_{\Sigma_V}(u), \quad Du = d.$$

Note that

$$\sqrt{1 + |d_i|^2} = \left|(d_i, 1)^T\right|$$

Idea: Introduce variable e with constraint $e_i = 1$ for all i.

$$\min_{u,d,e} \quad \underbrace{\sum_i \underbrace{\sqrt{e_i^2 + |d_i|^2}}_{=|(d_i,e_i)^T|}}_{=\|(d,e)\|_{2,1}} + \delta_{\Sigma_V}(u), \quad Du = d, e = \mathbf{1}$$

Single view 2.5D reconstruction

$$\min_{u,d,e} \quad \|(d,e)\|_{2,1} + \delta_{\Sigma_V}(u), \quad Du = d, e = \mathbf{1}$$

Now the proximity operators of the two functions are simple!

$$\min_{u,d,e} \max_{p,q} \|(d,e)\|_{2,1} + \delta_{\Sigma_V}(u) + \left\langle \begin{pmatrix} p \\ q \end{pmatrix}, \begin{pmatrix} -D & I & 0 \\ 0 & 0 & I \end{pmatrix} \begin{pmatrix} u \\ d \\ e \end{pmatrix} \right\rangle - \langle q, \mathbf{1} \rangle$$

Option 1: Use PDHG

$\rightarrow$ *Board*

Single view 2.5D reconstruction

Single view 2.5D reconstruction

Single view 2.5D reconstruction

Single view 2.5D reconstruction