Chapter 6

Stopping criteria, adaptivity, accelerations

Convex Optimization for Computer Vision
SS 2016

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

Customized proximal point algorithms

Structured optimization methods for

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

under the assumption of F and G being simple or - in the ADMM case - $(\partial G + \frac{1}{\tau} K^T K)^{-1}$ being easy to compute.

Goal: Find pair $(\hat{u}, \hat{p})$ with

$$-K^T \hat{p} \in \partial G(\hat{u}), \quad K\hat{u} \in \partial F^*(\hat{p})$$

Primal Dual-Hybrid Gradient (PDHG) method:

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

What is a good stopping criterion?

Michael Moeller
Thomas Möllenhoff
Emanuel Laude

Stopping customized proximal point algorithms

Generic form:

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

such that the matrix M is positive (semi-)definite.

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 $\partial F^*(p^{k+1})$?

We define the ***primal** and **dual** residuals*:

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

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

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

Note that this notion of convergences does not imply convergence of u^k and p^k yet!

Nevertheless, we know PDHG and ADMM do converge, and $\|r_d^{k+1}\|$ 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

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 $\partial F^*(p^{k+1})$.

More precisely:

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

Thus, we can use

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

to be scale-independent.

Michael Moeller
Thomas Möllenhoff
Emanuel Laude

Scaling the dual residuum

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 therefore scales with the magnitude of elements in $\partial G(u^{k+1})$.

More precisely:

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

Thus, we can use

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

to be 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, Douglas Rachford, PDHG, and any other "customized proximal point" algorithm 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

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

Note that z is exactly the "split" variable in the augmented Lagrangian based derivation of ADMM!

Michael Moeller
Thomas Möllenhoff
Emanuel Laude

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?

Michael Moeller
Thomas Möllenhoff
Emanuel Laude

Residual balancing

First option: Residual balancing! Let $(M_1, -K^T; -K, M_2)$ be positive definite. Pick τ^0 and σ^0 with $\tau^0\sigma^0 < 1$ as well as $\mu > 1$, $\alpha > 1$:

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

$$\tau^{k+1} = \frac{1}{\alpha}\tau^k, \quad \sigma^{k+1} = \alpha\sigma^k$$

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

$$\tau^{k+1} = \alpha\tau^k, \quad \sigma^{k+1} = \frac{1}{\alpha}\sigma^k$$

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

Why could this make sense?

Michael Moeller
Thomas Möllenhoff
Emanuel Laude

Unbalanced adaption

Second option: Fougner, Boyd '15: Let $(M_1, -K^T; -K, M_2)$ be positive definite. Pick τ^0 and σ^0 with $\tau^0\sigma^0 < 1$ as well as $\mu > 1$, $\alpha > 1$:

- If $\|r_d^k\| < \epsilon^{thresh}$ and $k > \mu k_1^{prev}$, do

$$\tau^{k+1} = \frac{1}{\alpha}\tau^k, \quad \sigma^{k+1} = \alpha\sigma^k, \quad k_1^{prev} \leftarrow k.$$

- If $\|r_p^k\| < \epsilon^{thresh}$ and $k > \mu k_2^{prev}$, do

$$\tau^{k+1} = \alpha\tau^k, \quad \sigma^{k+1} = \frac{1}{\alpha}\sigma^k, \quad k_2^{prev} \leftarrow k.$$

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

Convergence guarantees?

The previous two adaptive step size methods are heuristics that work well in practice.

In general, they have no convergence guarantees!

Common trick: Changing the parameters finitely many times only, reestablishes the convergence guarantees!

More appealing from a theoretical point of view: Decreasing the adaptivity of the stepsizes fast enough.

Michael Moeller
Thomas Möllenhoff
Emanuel Laude

Convergence guarantees with adaptive step sizes

Goldstein et al. 2015

Consider $M_1 = \frac{1}{\tilde{\tau}} I$, $M_2 = \frac{1}{\tilde{\sigma}} I$ with $\tilde{\sigma}\tilde{\tau} < \|K\|^{-2}$, and define

$$\delta^k = \min\left\{\frac{\tau^{k+1}}{\tau^k}, \frac{\sigma^{k+1}}{\sigma^k}, 1\right\}, \quad \phi^k = 1 - \delta^k$$

Let the following three conditions hold:

1. The sequences $\{\tau^k\}$, $\{\sigma^k\}$ remain bounded.
2. The sequence ϕ^k is summable.
3. It holds that $\tau^k \sigma^k < c < 1$.

Then the resulting adaptive PDHG algorithm converges.

Conjecture (for you to prove)

The same result holds for arbitrary M_1, M_2 provided that the matrix $(M_1, -K^T; -K, M_2)$ is positive definite.

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.

Convergence proof based on previous theorem.

Michael Moeller
Thomas Möllenhoff
Emanuel Laude

Sketch of proof

Sketch of the proof:

- The product $\tau^k \sigma^k$ does not change, thus 3. holds.
- It holds that

$$\phi^k = \begin{cases} 0 & \text{if stepsizes were not updated,} \\ \alpha^k & \text{if stepsizes were updated.} \end{cases}$$

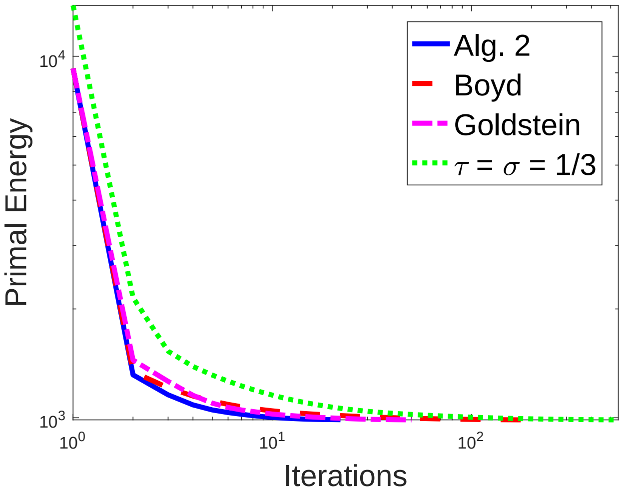
which means the j-th nonzero entry of $\{\phi^k\}$ is $(\alpha^0)^j$.

- $\sum_k \phi^k = \sum_{j \in I} (\alpha^0)^j < C$, thus condition 2 holds.
- Without restriction of generality we may drop those steps where the stepsize remained unchanged. We find

$$\tau^{j+1} \leq \frac{1}{1-\alpha^j} \tau^j \leq \left(\frac{1}{1-\alpha^j} \right)^j \tau^0 = \frac{1}{(1-\alpha^0 \beta^j)^j} \tau^0$$

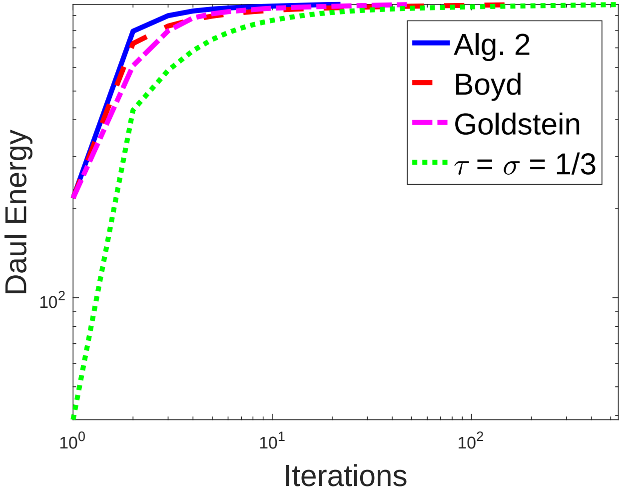
The factor $(1-\alpha^0 \beta^j)^j$ remains bounded from below and thus condition 1 follows. (For $x \geq -1$: $(1+x)^n \geq 1+nx$)

Example plot of convergence for ROF model



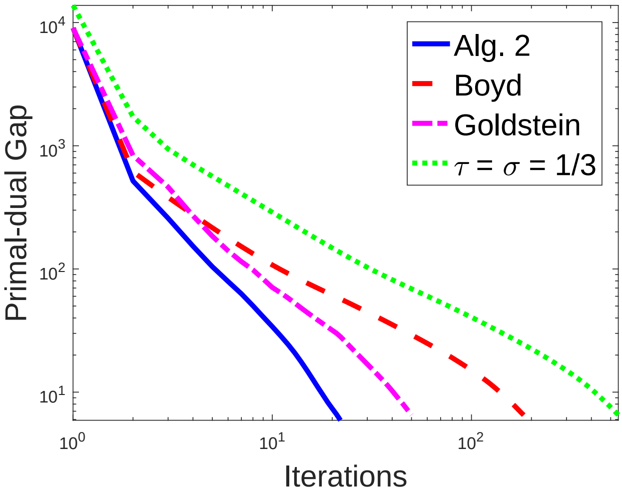
Michael Moeller
Thomas Möllenhoff
Emanuel Laude

Example plot of convergence for ROF model



Michael Moeller
Thomas Möllenhoff
Emanuel Laude

Example plot of convergence for ROF model



Backtracking

Condition 3 in the previous convergence result for adaptive stepsizes can also be weakened to

3. The saddle point problem

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

restricts either u or p to a bounded set. Furthermore there exists a constant c such that for all $k > 0$

$$\left\langle \begin{bmatrix} u^{k+1} - u^k \\ p^{k+1} - p^k \end{bmatrix}, \begin{bmatrix} \frac{1}{\tau^k} M_1 & -K^T \\ -K & \frac{1}{\sigma^k} M_2 \end{bmatrix} \begin{bmatrix} u^{k+1} - u^k \\ p^{k+1} - p^k \end{bmatrix} \right\rangle$$
$$\geq c \left\langle \begin{bmatrix} u^{k+1} - u^k \\ p^{k+1} - p^k \end{bmatrix}, \begin{bmatrix} \frac{1}{\tau^k} M_1 & 0 \\ 0 & \frac{1}{\sigma^k} M_2 \end{bmatrix} \begin{bmatrix} u^{k+1} - u^k \\ p^{k+1} - p^k \end{bmatrix} \right\rangle .$$

Under this condition the convergence result still holds.

Backtracking

The stability condition 3 from the previous slide can used to define a *backtracking* algorithm that works without knowing the constant $\|K\|^2$.

Define

$$b^k = \frac{2\tilde{\tau}\tilde{\sigma}\tau^k\sigma^k\langle p^{k+1} - p^k, K(u^{k+1} - u^k)\rangle}{\gamma\tilde{\sigma}\sigma^k\|u^{k+1} - u^k\|^2 + \gamma\tilde{\tau}\tau^k\|p^{k+1} - p^k\|^2}$$

for some $\gamma \in]0,1[$.

If $b^k \leq 1$ keep iterating, if $b^k > 1$ update

$$\tau^{k+1} = \beta\tau^k/b^k, \quad \sigma^{k+1} = \beta\sigma^k/b^k$$

for $\beta \in]0,1[$.

Key insight to prove convergence: $b^k > 1$ can only happen finitely many times.

Michael Moeller
Thomas Möllenhoff
Emanuel Laude

Preconditioning

Generic customized proximal point algorithms:

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

overrelaxation on primal variable u

or

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

overrelaxation on dual variable p.

What choices can we make beyond $M_1 = \frac{1}{\tau} I$ and $M_2 = \frac{1}{\sigma} I$?

Michael Moeller
Thomas Möllenhoff
Emanuel Laude

Recalling some customized proximal point algorithms

A computation on the board shows

- **Primal ADMM, u update first**

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

- **Corresponding dual ADMM**

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

- **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**

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

Recalling some customized proximal point algorithms

Practical experience:
ADMM makes more progress per iteration than PDHG!

Interpretation: PDHG approximates $\lambda K^T K$ by $\frac{1}{\sigma} I$.
$\rightarrow$ Crude approximation!

Idea: Use

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

for matrices M_1, M_2 that introduce more knowledge about K!

To avoid the difficult resolvents: Use diagonal M_1 and M_2!

Preconditioning matrices

Pock, Chambolle 2011

Let M_1 and M_2 be symmetric positive definite maps that satisfy

$$\left\| M_2^{-1/2} K M_1^{-1/2} \right\|^2 < 1,$$

then the matrix

$$M := \begin{bmatrix} M_1 & -K^T \\ -K & M_2 \end{bmatrix}$$

is positive definite.

Preconditioning matrices

Pock, Chambolle 2011

Let $M_1 = \mathrm{diag}(m_j^1)$ and $M_2 = \mathrm{diag}(m_i^2)$ with

$$m_j^1 = \sum_i |K_{i,j}|^{2-\alpha}, \qquad m_i^2 = \sum_j |K_{i,j}|^{\alpha}.$$

Then

$$\|M_2^{-1/2} K M_1^{-1/2}\|^2 \leq 1$$

holds for all $\alpha \in [0, 2]$.

The above theorem provides an easy way of determining diagonal preconditioners!

To get a strict inequality we can multiply M_1 and/or M_2 by any factor > 1.

When does preconditioning make sense?

For $K \approx \nabla$ this preconditioning has (almost) no effect!

If the row sums or the column sums of K differ a lot, then the previous preconditioning has a strong effect:

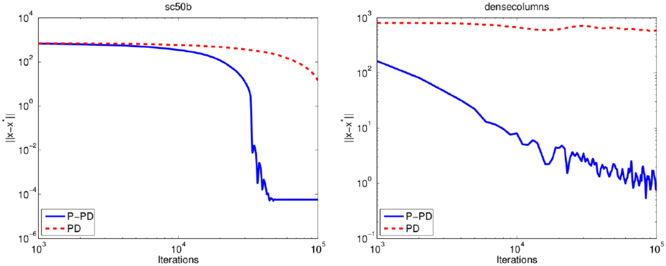


Figure 1. On problems with irregular structure, the proposed preconditioned algorithm (P-PD) converges significantly faster than the algorithm of [5] (PD).

From Pock, Chambolle 2011: *Diagonal preconditioning for first order primal-dual algorithms in convex optimization.*

Open research questions

- Convergence for triangular matrices M_1, M_2?
- Symmetric non-diagonal preconditioner which still allow efficient solutions?
- Iterative adaptation of diagonal preconditioners?
- Why does ”preconditioning” help beyond diagonal matrices?
- Convergence estimates that reveal influence of preconditioners?

Gaining some momentum

Assume we look for a $\hat{x}$ such that

$$0 \in (A + B)\hat{x}$$

for maximally monotone operators A and B. Then the scheme

$$y^k = x^k + \alpha_k(x^k - x^{k-1}),$$
$$x^{k+1} = (M + \lambda_k A)^{-1}(M - \lambda_k B)(y^k),$$

converges for certain choices of extrapolation parameter α_k, positive (semi-)definite matrix M, and stepsize λ_k.

Example 1: $A = 0$, $M = I$, $\lambda_k = \tau$, $B = \nabla E$:

$$x^{k+1} = y^k - \nabla E(y^k)$$

Example 2: $A = \partial E_1$, $M = I$, $\lambda_k = \tau$, $B = \nabla E_2$:

$$x^{k+1} = \text{prox}_{\tau E_1}(y^k - \nabla E_2(y^k))$$

These methods are optimal in the sense of Nesterov. Similar techniques may also accelerate primal-dual algorithms.[1]

[1] Details: "An inertial forward-backward algorithm for monotone inclusions".

Very brief desired topic:
ADMM and its convergence

Possible convergence analysis of ADMM on 1 slide

- The subdifferential ∂E is a maximally monotone operator.
- The resolvent $(I + A)^{-1}$ of a maximally monotone operator A is non-expansive.
- An operator F is called averaged, if $F = \theta T + (1 - \theta) I$ for $\theta \in]0, 1[$ and T being non-expansive.
- If F is an averaged operator, the fixed point iteration $x^{k+1} = F x^k$ converges to some x^* with $x^* = F x^*$.
- Specially structured problem: $0 \in \partial G(u) + K^T \partial F(Ku)$, i.e. $0 \in A(u) + B(u)$, for A and B being maximally monotone.
- For $C_A = 2(I + A)^{-1} - I$ denoting the Cayley operator, $C_A C_B$ is a non-expansive operator!
- DRS/ADMM: Do a fixed point iteration with

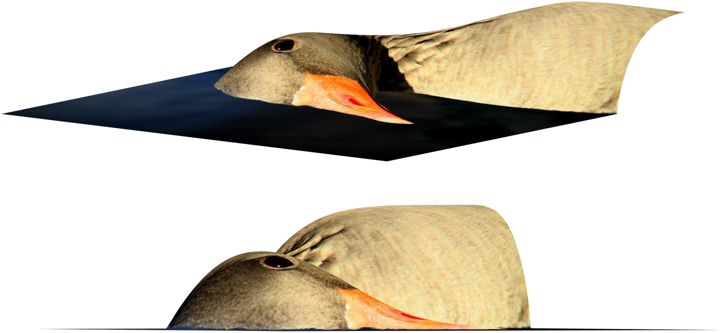
$$F = \frac{1}{2} I + \frac{1}{2} C_A C_B,$$

and $u^* = (I + B)^{-1} x^*$ will meet $0 \in A(u^*) + B(u^*)$.

Case study:
Single view 3d reconstruction

Remember 2.5d reconstruction?

Remember 2.5d reconstruction?



Remember 2.5d reconstruction?

Assumptions:

- Minimal surface
- User-defined volume
- No volume outside of the contour

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

Michael Moeller
Thomas Möllenhoff
Emanuel Laude

What about full 3d reconstruction?

First question: How do we represent the surface/volume?

Common technique: $u : \Omega \times \mathbb{R} \to \{0, 1\}$.

- $u(x) = 1$ means this voxel is occupied with the object
- $u(x) = 0$ means there is no object at this pixel

Via the celebrated co-area formula, one may show that if

$$u(x) = \begin{cases} 1 & \text{if } x \in M, \\ 0 & \text{otherwise,} \end{cases}$$

one has

$$TV(u) := \int \|Du\| = \text{Area}(M).$$

For details see e.g. Chambolle et at. 2009, *An introduction to Total Variation for Image Analysis*.

Michael Moeller
Thomas Möllenhoff
Emanuel Laude

Single image 3D reconstruction

What is a reasonable first model for single image 3D reconstruction given a silhouette S?

- $\int_{\Omega \times \mathbb{R}} u(x)\, dx = V$, where V is a user given volume.
- Constraint: $u(x_1, x_2, z) = 0\ \forall z$ if $(x_1, x_2) \notin S$.
- Constraint: $u(x_1, x_2, 0) = 1\ \forall z$ if $(x_1, x_2) \in S$.
- Find the minimal surface, i.e. minimize

$$\int_{\Omega \times \mathbb{R}} |\nabla u(x)|\, dx$$

- Or with prior information possibly use a weighting

$$\int_{\Omega \times \mathbb{R}} g(x) |\nabla u(x)|\, dx$$

- Since the constraint $u(x) \in \{0, 1\}$ is difficult, use $u \in [0, 1]$.

Single image 3D reconstruction

Discretized model:

$$\min_{u\in\mathbb{R}^{n_y\times n_x\times n_d}} \sum_{i,j,k} g_{i,j,k}\cdot |(Du)_{i,j,k}| + \delta_{\Sigma_V}(u) + \delta_{u_{i,j,n_d/2}=1}(u) + \delta_{[0,1]}(u)$$

Let us apply the PDHG method to minimize the above energy.

First step: Bring the energy into a suitable saddle point form

Note that one can absorb the $g_{i,j,k}$ into D by multiplying D with a diagonal matrix with the $g_{i,j,k}$ on the diagonal.

Michael Moeller
Thomas Möllenhoff
Emanuel Laude

Single image 3D reconstruction

Discretized model:

$$\min_{u \in \mathbb{R}^{n_y \times n_x \times n_d}} \|Du\|_{2,1} + \delta_{\Sigma_V}(u) + \delta_{u_{i,j,n_d/2}=1}(u) + \delta_{[0,1]}(u)$$

As usual we reformulate

$$\min_u \max_p \ \langle p, Du \rangle - \delta_{\|\cdot\|_{2,\infty}}(p) + \delta_{\Sigma_V}(u) + \delta_{u_{i,j,n_d/2}=1}(u) + \delta_{[0,1]}(u)$$

We know the $\delta_{\|\cdot\|_{2,\infty}}(p)$ prox is easy. What about the remaining problem in u?

At least as difficult as a non-negative ℓ^1 projection[2], but with an additional bound on each component of u! We should simplify further!

[2] See Duchi et al. 2008

Single image 3D reconstruction

$$\min_u \max_p \; \langle p, Du \rangle - \delta_{\|\cdot\|_{2,\infty}}(p) + \delta_{\Sigma_V}(u) + \delta_{u_{i,j,n_d/2}=1}(u) + \delta_{[0,1]}(u)$$

One could ”dualize” either $\delta_{[0,1]}(u)$ or $\delta_{\Sigma_V}(u)$ and the remaining prox in u would be easy!

Lets use

$$\delta_{\Sigma_V}(u) = \sup_{q \in \mathbb{R}} q(\mathbf{1}^T u - V)$$

Michael Moeller
Thomas Möllenhoff
Emanuel Laude

Single image 3D reconstruction

We arrive at

$$\min_u \max_{p,q} \ \langle p, Du\rangle - \delta_{\|\cdot\|_{2,\infty}\leq 1}(p) + q(\mathbf{1}^T u - V) + \delta_{u_{i,j,n_d/2}=1}(u) + \delta_{[0,1]}(u)$$

or equivalently

$$\min_u \max_{p,q} \ \delta_{u_{i,j,n_d/2}=1}(u) + \delta_{[0,1]}(u) + \left\langle \begin{pmatrix} p \\ q \end{pmatrix}, \begin{pmatrix} D \\ \mathbf{1}^T \end{pmatrix} u \right\rangle - qV - \delta_{\|\cdot\|_{2,\infty}\leq 1}(p)$$

Let's apply (PDHG)!

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

Michael Moeller
Thomas Möllenhoff
Emanuel Laude

Single image 3D reconstruction

$$p^{k+1} = \text{prox}_{\delta_{\|\cdot\|_{2,\infty}\leq 1}}(p^k + \sigma D\bar{u}^k)$$

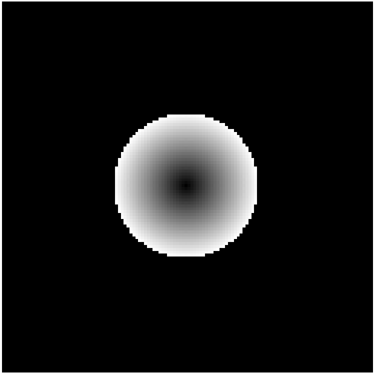
$$q^{k+1} = \operatorname*{argmin}_q 0.5(q - q^k - \sigma \mathbf{1}^T \bar{u}^k)^2 + \sigma V q$$
$$= q^k + \sigma(\mathbf{1}^T u^k - V)$$

$$u^{k+1}_{i,j,l} = \begin{cases} 1 & \text{if } l = n_d/2 \\ \max(0, \min(1, u^{k+1}_{i,j,l} - \tau(D^* p^{k+1} + q^{k+1}\,\mathbf{1}))) & \text{else} \end{cases}$$

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

Single image 3D reconstruction

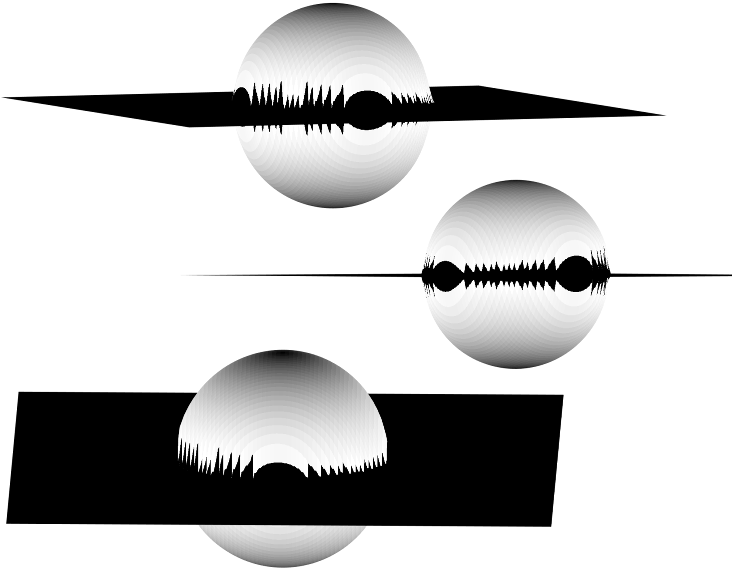
Sanity check:



with Volume $= \frac{4}{3}\pi \text{radius}^3$.

Michael Moeller
Thomas Möllenhoff
Emanuel Laude

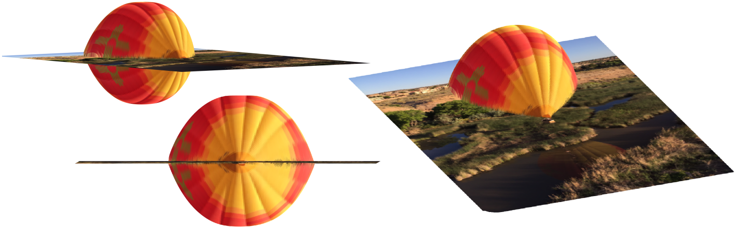
Single image 3D reconstruction



Single image 3D reconstruction

Michael Moeller
Thomas Möllenhoff
Emanuel Laude

Single image 3D reconstruction

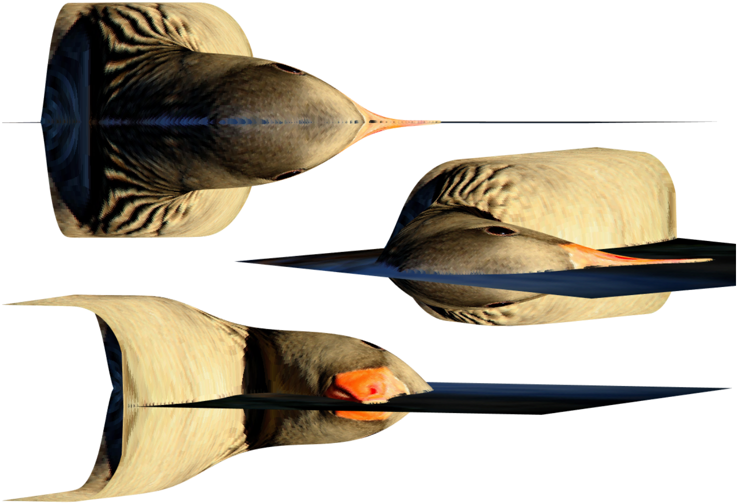


Nice results (except some remaining discretization artifacts).

Michael Moeller
Thomas Möllenhoff
Emanuel Laude

Single image 3D reconstruction

Single image 3D reconstruction



Results (with a restriction on the maximal thickness).