Decomposition by operator-splitting methods and applications in image deblurring

Daniel O'Connor (Department of Mathematics, UCLA)

Lieven Vandenberghe (Electrical Engineering, UCLA)

Workshop on Optimization for Modern Computation

BICMR, Beijing, September 2–4, 2014

Primal-dual decomposition

$$\text{minimize} \quad f(x) + g(Ax)$$

- f, g are 'simple' convex functions (indicators of simple sets, norms, . . .)
- A is a structured matrix
- widely used format in literature on multiplier and splitting algorithms

This talk: decomposition by splitting primal-dual optimality conditions

$$0 \in \begin{bmatrix} 0 & A^T \\ -A & 0 \end{bmatrix} \begin{bmatrix} x \\ z \end{bmatrix} + \begin{bmatrix} \partial f(x) \\ \partial g^*(z) \end{bmatrix}$$

and applications in image deblurring

Models of image blur

$$b = Kx + w$$

- x is exact image, Kx is image blurred by blurring operator K
- b is observed image, equal to blurred image plus noise w

Space-invariant blur: structure of K depends on boundary conditions

- periodic: WKW^H is diagonal where W is 2D DFT matrix
- zero/replicate: $K = K_{\mathrm{c}} + K_{\mathrm{s}}$ with K_{c} diagonalizable by DFT, K_{s} sparse

Space-varying blur (Nagy-O'Leary model): $K = \sum_{i=1}^{m} U_i K_i$

- K_i: space-invariant blurring operators
- U_i: positive diagonal matrices that sum to identity

Deblurring via convex optimization

$$\text{minimize} \quad \phi_{\mathrm{f}}(Kx - b) + \phi_{\mathrm{s}}(Dx) + \phi_{\mathrm{r}}(x)$$

Data fidelity term ϕ_{f}

- convex penalty, $e.g.$, squared 2-norm, 1-norm, Huber penalty, . . .
- indicator for convex set, $e.g.$, 2-norm ball

Smoothing term ϕ_{s}

- D is discretized first derivative, or a wavelet/shearlet transform matrix
- ϕ_{s} is a norm, $e.g.$, for total variation reconstruction, a sum of 2-norms

$$\phi_{\mathrm{s}}(u, v) = \gamma \|(u, v)\|_{\mathrm{iso}} = \gamma \sum_{i=1}^{n} \sqrt{u_i^2 + v_i^2}$$

Deblurring via convex optimization

$$\text{minimize} \quad \phi_{\mathrm{f}}(Kx - b) + \phi_{\mathrm{s}}(Dx) + \phi_{\mathrm{r}}(x)$$

Regularization term ϕ_{r}

- penalty on x
- indicator for convex set, for example, $\{x \mid 0 \leq x \leq \mathbf{1}\}$

In composite form: minimize $f(x) + g(Ax)$ with

$$f(x) = \phi_{\mathrm{r}}(x), \qquad A = \begin{bmatrix} K \\ D \end{bmatrix}, \qquad g(u, v) = \phi_{\mathrm{f}}(u - b) + \phi_{\mathrm{s}}(v)$$

Outline

- Introduction
- **Douglas-Rachford splitting method**
- Primal-dual splitting
- Space-varying blur

Monotone operator

A set-valued mapping $\mathcal{F}$ is **monotone** if

$$(u-v)^T(y-x) \geq 0, \qquad \forall x, y \in \mathbf{dom}\,\mathcal{F},\ u \in \mathcal{F}(x),\ v \in \mathcal{F}(y)$$

- subdifferential ∂f of closed convex function f
- skew-symmetric linear operator, for example,

$$\mathcal{F}(x,z) = \begin{bmatrix} 0 & A^T \\ -A & 0 \end{bmatrix} \begin{bmatrix} x \\ z \end{bmatrix}$$

- sums of monotone operators, for example,

$$\mathcal{F}(x,z) = \begin{bmatrix} 0 & A^T \\ -A & 0 \end{bmatrix} \begin{bmatrix} x \\ z \end{bmatrix} + \begin{bmatrix} \partial f(x) \\ \partial g^*(z) \end{bmatrix}$$

Resolvent

The **resolvent** of a monotone operator $\mathcal{F}$ is the operator

$$(I + t\mathcal{F})^{-1} \qquad \text{(with } t > 0\text{)}$$

Properties (for maximal monotone $\mathcal{F}$)

- $y = (I + t\mathcal{F})^{-1}(x)$ exists and is unique for all x
- y is the (unique) solution of the inclusion problem $x \in y + t\mathcal{F}(y)$

Examples

- resolvent of subdifferential ∂f is called proximal operator of f
- for linear monotone $\mathcal{F}$, resolvent $(I + t\mathcal{F})^{-1}$ is matrix inverse

Proximal operator

The **proximal operator** of a closed convex function f is the mapping

$$\mathrm{prox}_{tf}(x) = \underset{y}{\mathrm{argmin}} \left(f(y) + \frac{1}{2t}\|y - x\|_2^2 \right)$$

Examples

- $f(x) = \delta_C(x)$ (indicator of closed convex set C): Euclidean projection

$$\mathrm{prox}_{tf}(x) = P_C(x) = \underset{y}{\mathrm{argmin}}\, \|y - x\|_2^2$$

- $f(x) = \|x\|$: shrinkage operation

$$\mathrm{prox}_{tf}(x) = x - P_{tC}(x), \qquad C \text{ is unit ball for dual norm}$$

Calculus rules for proximal operators

Separable function: if $f(x_1, x_2) = f_1(x_1) + f_2(x_2)$, then

$$\mathrm{prox}_f(x_1, x_2) = \big(\mathrm{prox}_{f_1}(x_1), \mathrm{prox}_{f_2}(x_2)\big)$$

Moreau decomposition: relates prox-operators of conjugates

$$\mathrm{prox}_{tf^*}(x) + t\mathrm{prox}_{t^{-1}f}(x/t) = x$$

Composition with affine mappig: $f(x) = g(Ax + b)$ with $AA^T = aI$

$$\mathrm{prox}_f(x) = (I - \frac{1}{a}A^T A)x + \frac{1}{a}A^T \big(\mathrm{prox}_{ag}(Ax + b) - b\big)$$

Douglas-Rachford splitting

Problem: given maximal monotone operators $\mathcal{A}$, $\mathcal{B}$, solve

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

Algorithm (Lions & Mercier, 1979)

$$\begin{aligned} x^+ &= (I + t\mathcal{A})^{-1}(z) \\ y^+ &= (I + t\mathcal{B})^{-1}(2x^+ - z) \\ z^+ &= z + \rho(y^+ - x^+) \end{aligned}$$

- x converges under weak conditions (for any $t > 0$ and $\rho \in (0, 2)$)
- useful when resolvents of $\mathcal{A}$, $\mathcal{B}$ are inexpensive, but not resolvent of sum
- includes other well-known algorithms as special cases ($e.g.$, ADMM)

Outline

- Introduction
- Douglas-Rachford splitting method
- **Primal-dual splitting**
- Space-varying blur

Primal-dual splitting

Composite problem and dual

$$\text{minimize} \quad f(x) + g(Ax) \qquad\qquad \text{maximize} \quad -f^*(-A^T z) - g^*(z)$$

Primal-dual optimality conditions

$$0 \in \underbrace{\begin{bmatrix} \partial f(x) \\ \partial g^*(z) \end{bmatrix}}_{\mathcal{A}(x,z)} + \underbrace{\begin{bmatrix} 0 & A^T \\ -A & 0 \end{bmatrix} \begin{bmatrix} x \\ z \end{bmatrix}}_{\mathcal{B}(x,z)}$$

Resolvent computations

- $\mathcal{A}$: prox-operators of f and g
- $\mathcal{B}$: solution of a linear equation with coefficient $I + t^2 A^T A$

Example: constrained L1-TV deblurring

$$\begin{array}{ll} \text{minimize} & \|Kx - b\|_1 + \gamma \|Dx\|_{\text{iso}} \\ \text{subject to} & 0 \leq x \leq \mathbf{1} \end{array}$$

- Gaussian blur with salt-and-pepper noise; periodic boundary conditions
- $I + K^T K + D^T D$ diagonalizable by DFT
- 1024×1024 image


original


blurred


restored

Primal Douglas-Rachford splitting

Equivalent problem (δ is indicator function of $\{0\}$)

$$\text{minimize} \quad f(x) + g(Ax) \quad \longrightarrow \quad \text{minimize} \; \underbrace{f(x) + g(y)}_{F(x,y)} + \underbrace{\delta(Ax - y)}_{G(x,y)}$$

Algorithm: Douglas-Rachford splitting applied to optimality conditions

$$0 \in \partial F(x, y) + \partial G(x, y)$$

Resolvent computations

- ∂F requires prox-operators of f, g
- ∂G requires linear equation with coefficient $I + A^T A$

hence, similar complexity per iteration as primal-dual splitting

Alternating direction method of multipliers (ADMM)

Douglas-Rachford applied to dual, after introducing splitting variable u

$$\text{minimize} \quad f(x) + g(Ax) \quad \longrightarrow \quad \begin{array}{ll} \text{minimize} & f(u) + g(y) \\ \text{subject to} & \begin{bmatrix} I \\ A \end{bmatrix} x - \begin{bmatrix} u \\ y \end{bmatrix} = 0 \end{array}$$

ADMM: alternating minimization of augmented Lagrangian

$$f(u) + g(y) + w^T(x-u) + z^T(Ax-y) + \frac{t}{2}\left(\|x-u\|_2^2 + \|Ax-y\|_2^2\right)$$

- minimization over x: linear equation with coefficient $I + A^T A$
- minimization over (u, y): prox-operators of f, g

hence, similar complexity per iteration as primal-dual splitting

Chambolle-Pock method

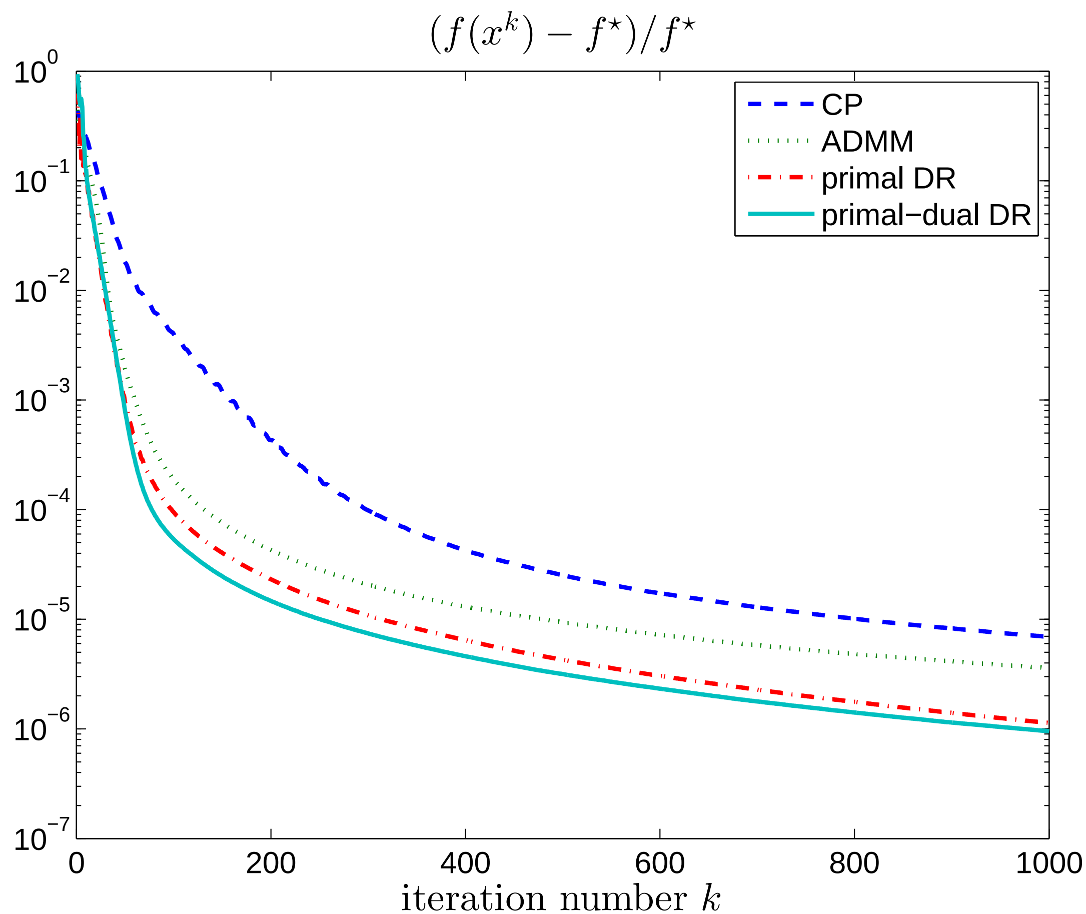
$$0 \in \begin{bmatrix} \partial f(x) \\ \partial g^*(z) \end{bmatrix} + \begin{bmatrix} 0 & A^T \\ -A & 0 \end{bmatrix} \begin{bmatrix} x \\ z \end{bmatrix}$$

Algorithm

$$\begin{aligned} z^+ &= \operatorname{prox}_{tg^*}(z + tA\bar{x}^+) \\ x^+ &= \operatorname{prox}_{sf}(x - sA^T z^+) \\ \bar{x}^+ &= 2x^+ - x \end{aligned}$$

- convergence requires $\sqrt{st} < 1/\|A\|_2$
- no linear equations with A; only multiplications with A and A^T

Convergence



~ 1.4 seconds per iteration for each method

Additive structure in A

$$\text{minimize} \quad f(x) + g(Ax) \qquad\qquad \text{maximize} \quad -f^*(-A^T z) - g^*(z)$$

- f, g have inexpensive prox-operators
- $A = B + C$ with structured B and C: equations with coefficients

$$I + B^T B, \qquad I + C^T C$$

are easy to solve, but not $I + A^T A$

Extended primal-dual optimality conditions

$$0 \in \begin{bmatrix} 0 \\ \partial g(y) \\ 0 \\ \partial f^*(w) \end{bmatrix} + \begin{bmatrix} 0 & 0 & A^T & I \\ 0 & 0 & -I & 0 \\ -A & I & 0 & 0 \\ -I & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix}$$

Primal-dual splitting

$$0 \in \underbrace{\begin{bmatrix} 0 \\ \partial g(y) \\ 0 \\ \partial f^*(w) \end{bmatrix} + \begin{bmatrix} 0 & 0 & B^T & 0 \\ 0 & 0 & 0 & 0 \\ -B & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix}}_{\mathcal{A}(x,y,z,w)}$$

$$+ \underbrace{\begin{bmatrix} 0 & 0 & C^T & I \\ 0 & 0 & -I & 0 \\ -C & I & 0 & 0 \\ -I & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix}}_{\mathcal{B}(x,y,z,w)}$$

Resolvent computations

- $\mathcal{A}$: prox-operators of f, g, linear equation $I + t^2 B^T B$
- $\mathcal{B}$: linear equation with coefficient $I + \frac{t^2}{(1+t^2)^2} C^T C$

TV-L1 deblurring with replicate boundary conditions

$$\begin{array}{ll} \text{minimize} & \|(K_{\mathrm{c}} + K_{\mathrm{s}})x - b\|_1 + \gamma \|(D_{\mathrm{c}} + D_{\mathrm{s}})x\|_{\mathrm{iso}} \\ \text{subject to} & 0 \preceq x \preceq \mathbf{1} \end{array}$$

- K_{c}, D_{c}: operators for periodic boundary conditions
- K_{s}, D_{s}: sparse correction for replicate boundary conditions

blurry, noisy image

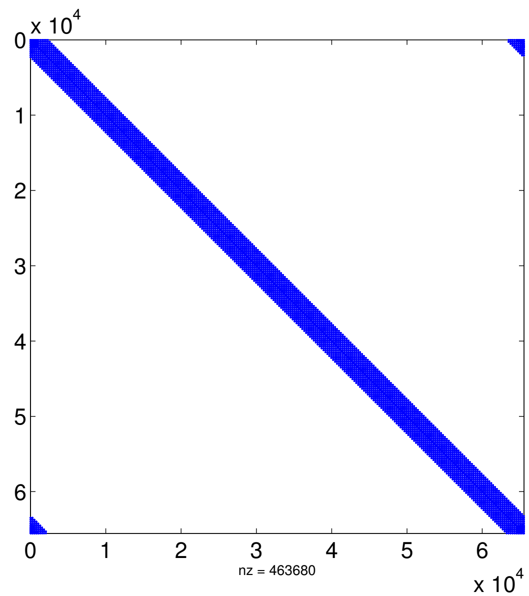
deblurred using periodic b.c.

deblurred using replicate b.c.

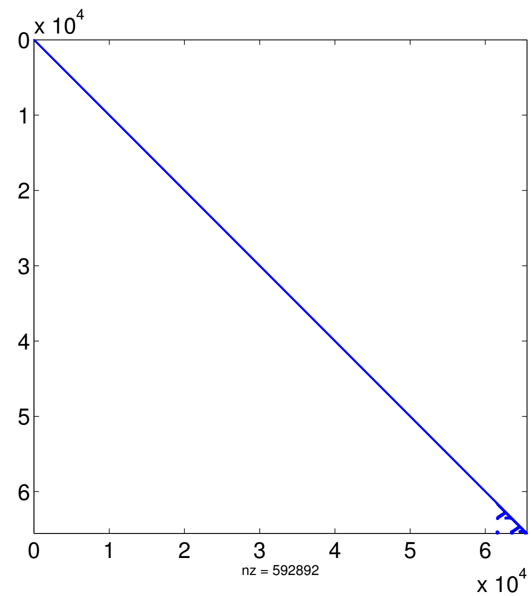
Handling replicate boundary conditions

$$K = K_{\rm c} + K_{\rm s}, \qquad D = D_{\rm c} + D_{\rm s}$$

- $K_{\rm c}$, $D_{\rm c}$: operators assuming periodic boundary conditions
- $I + K_{\rm c}^T K_{\rm c} + D_{\rm c}^T D_{\rm c}$ is diagonalized by DFT
- $E = I + K_{\rm s}^T K_{\rm s} + D_{\rm s}^T D_{\rm s}$ is sparse



pattern of E



Cholesky factor of E

Primal Douglas-Rachford splitting

Equivalent problem: introduce splitting variables $\tilde{x}$, $\tilde{y}$

$$\text{minimize} \quad \underbrace{f(x) + g(y + \tilde{y}) + \delta(x - \tilde{x})}_{F(x,\tilde{x},y,\tilde{y})} + \underbrace{\delta(Bx - y) + \delta(C\tilde{x} - \tilde{y})}_{G(x,\tilde{x},y,\tilde{y})}$$

and apply Douglas-Rachford method to find zero of

$$0 \in \partial F(x, \tilde{x}, y, \tilde{y}) + \partial G(x, \tilde{x}, y, \tilde{y})$$

Resolvent computations

- ∂F: require prox-operators of f, g
- ∂G: linear equations with coefficients $I + B^T B$, $I + C^T C$

more variables but similar complexity per iteration as primal-dual splitting

ADMM

Equivalent problem: introduce another splitting variable u

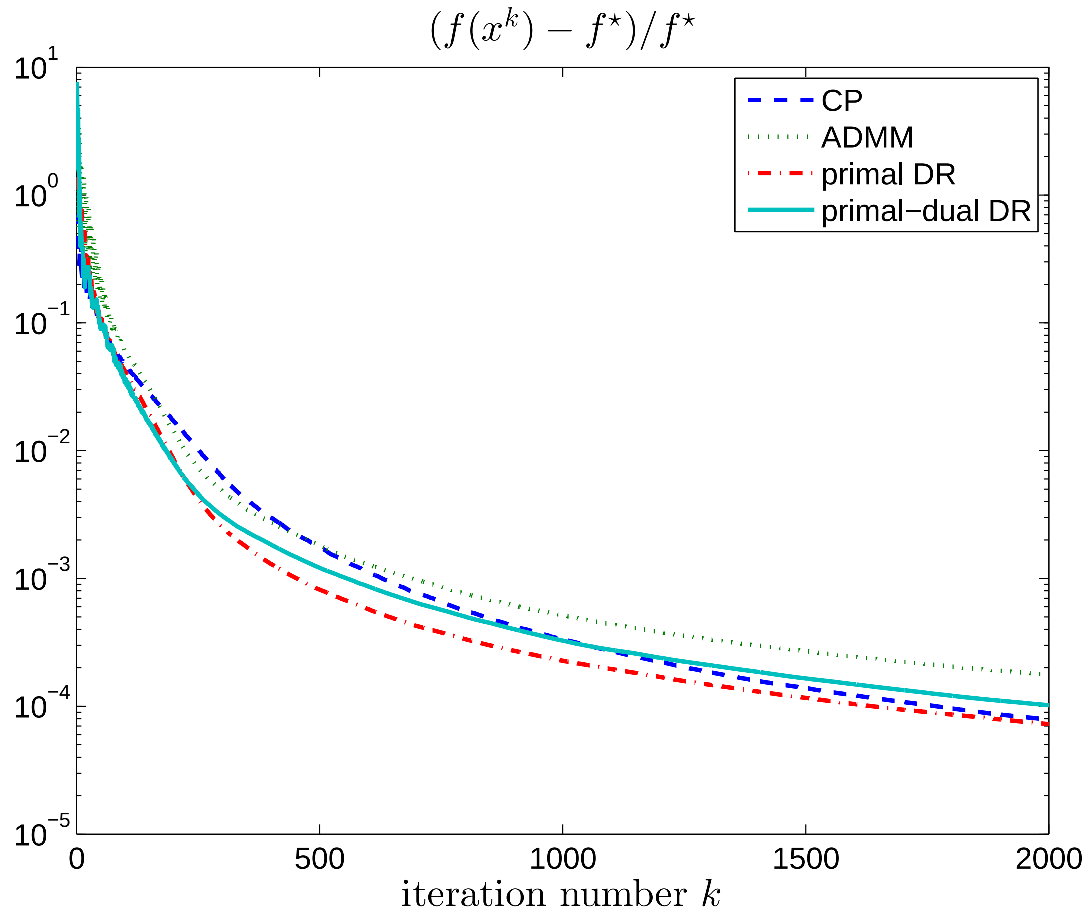
$$\begin{array}{ll} \text{minimize} & f(u) + g(y + \tilde{y}) \\ \text{subject to} & \begin{bmatrix} I & 0 \\ 0 & I \\ B & 0 \\ 0 & C \end{bmatrix} \begin{bmatrix} x \\ \tilde{x} \end{bmatrix} - \begin{bmatrix} I & 0 & 0 \\ I & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & I \end{bmatrix} \begin{bmatrix} u \\ y \\ \tilde{y} \end{bmatrix} = 0 \end{array}$$

ADMM: alternating minimization of augmented Lagrangian requires

- linear equations with coefficients $I + B^T B$, $I + C^T C$
- prox-operators of f, g

even more variables, but same complexity per iteration

Convergence



~ 0.1 seconds per iteration

Domain decomposition

- divide image in rectangular regions
- blurring and derivative operators are block-diagonal plus sparse
- diagonal blocks are operations on regions, with periodic boundary conds.

Example: constrained TV-L1 deblurring on 256×256 image

blurry, noisy image

after 5 iterations

restored image

Tight-frame regularization

$$\text{minimize} \quad \frac{1}{2}\|Kx - b\|_2^2 + \gamma\|Dx\|_1$$

- $K = K_{\mathrm{c}} + K_{\mathrm{s}}$ for blurring with replicate boundary conditions
- D is shearlet tight frame: satisfies $D^T D = \alpha I$
- 256×256 image

noisy, blurred

restored

Outline

- Introduction
- Douglas-Rachford splitting method
- Primal-dual splitting
- **Space-varying blur**

Space-varying blurring

Blurring model (Nagy and O'Leary, 1998)

$$K = U_1K_1 + U_2K_2 + \cdots + U_mK_m$$

- K_i are blurring matrices for space invariant kernels
- U_i are positive diagonal matrices with $U_1 + \cdots + U_m = I$
- K is not diagonalizable by DFT or DCT

Convex deblurring problem

$$\text{minimize} \quad \phi_{\mathrm{f}}(Kx - b) + \phi_{\mathrm{s}}(Dx) + \phi_{\mathrm{r}}(x)$$

- we assume ϕ_{f} is **separable** ($e.g.$, squared Euclidean norm, L1-norm)
- D is tight frame or derivative operator

Splitting method

$$\text{minimize} \quad \phi_{\mathrm{f}}(Kx-b)+\phi_{\mathrm{s}}(Dx)+\phi_{\mathrm{r}}(x)$$

As composite problem: minimize $f(x)+g(Ax)$ with $f(x)=\phi_{\mathrm{r}}(x)$,

$$\begin{array}{rcl} g(y_1,\ldots,y_{m+1}) & = & \phi_{\mathrm{f}}(U_1y_1+\cdots+U_my_m)+\phi_{\mathrm{s}}(y_{m+1}) \\ A & = & \left[\begin{array}{cccc} K_1^T & \cdots & K_m^T & D^T \end{array}\right]^T \end{array}$$

- $I+A^TA$ is diagonalizable by a DFT, hence easy to invert:

$$I+A^TA=I+\sum_{i=1}^{m}K_i^TK_i+D^TD$$

- prox-operators of f and g reduce to prox-operators of ϕ_{r}, ϕ_{f}, ϕ_{s}

Example

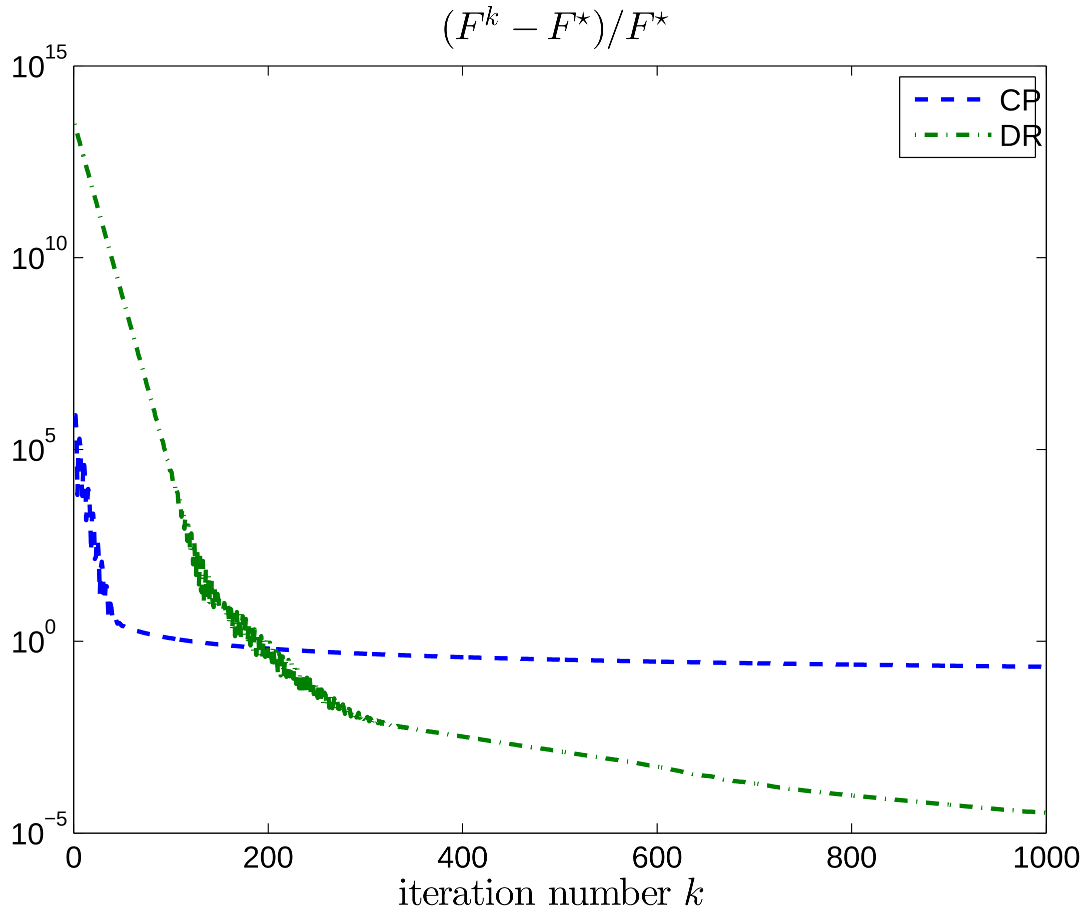
- 512×512 output image, 528×528 input (free boundary conditions)
- $m = 4$ kernels (one for each quadrant of the image)

noisy, blurred

L2-TV deblurred

~ 0.2 seconds per iteration (cost of a small number of FFTs)

Convergence



~ 0.2 seconds per iteration (cost of a small number of FFTs)

Motion deblurring

Example and software from Chakrabarti, Zickler, Freeman (2010)

- 367×600 image
- algorithm estimates motion blur kernel and segments out blurred region

image with motion blur

restored image

- segmentation used to build Nagy-O'Leary model with two kernels
- L2-TV deblurring using primal-dual Douglas-Rachford method

Summary

Douglas-Rachford splitting applied to primal-dual optimality conditions of

$$\text{minimize} \quad f(x) + g(Ax)$$

- f and g have inexpensive prox-operators
- A is structured: $I + A^T A$ is easy to invert
- extension: $A = B + C$ with $I + B^T B$, $I + C^T C$ easy to invert
- applications in image deblurring
- extends primal-dual decomposition (f, g separable, A angular + sparse)