ADAPTIVITY TO IMPROVE CONVERGENCE

Image and Signal Processing Seminars

Adriana González

ISPGroup - UCL/ICTEAM

February 19, 2014

Outline

1. General optimization problems
2. Iterative algorithms
3. Step-size selection using ski-rental problem
4. Step-size selection using residuals balancing
5. Numerical examples

Outline

1. General optimization problems
2. Iterative algorithms
3. Step-size selection using ski-rental problem
4. Step-size selection using residuals balancing
5. Numerical examples

General optimization problems

$$\min_{\mathbf{x}\in\mathcal{X}} \; F(\mathbf{K}\mathbf{x}) + G(\mathbf{x})$$

- Examples
  - Denoising

$$\min_{\mathbf{x}} \psi(\mathbf{x}) \text{ s.t. } \|\mathbf{x}-\mathbf{y}\|_2 \le \varepsilon; \qquad \min_{\mathbf{x}} \psi(\mathbf{x}) \text{ s.t. } \|\mathbf{x}-\mathbf{y}\|_1 \le \varepsilon$$

  - Deconvolution

$$\min_{\mathbf{x}} \psi(\mathbf{x}) \text{ s.t. } \|\mathbf{h} \star \mathbf{x} - \mathbf{y}\|_2 \le \varepsilon$$

  - ODT [1]

$$\min_{\mathbf{x}} \|\mathbf{x}\|_{\mathrm{TV}} \text{ s.t. } \|\boldsymbol{\Phi}\mathbf{x}-\mathbf{y}\|_2 \le \varepsilon, \mathbf{x} \succeq 0, \mathbf{x}_\Omega = 0$$

[1] A. González, L. Jacques, C. De Vleeschouwer, and P. Antoine, Compressive optical deflectometric tomography: A constrained total-variation minimization approach. To appear in Inverse Problems and Imaging Journal, 2014, preprint, arXiv:1209.0654.

Outline

1. General optimization problems
2. Iterative algorithms
3. Step-size selection using ski-rental problem
4. Step-size selection using residuals balancing
5. Numerical examples

Iterative algorithms

$$\min_{\mathbf{x}\in\mathcal{X}} \; F(\mathbf{x}) + G(\mathbf{K}\mathbf{x})$$

$$\min_{\mathbf{x}\in\mathcal{X};\ \mathbf{z}\in\mathcal{Z}} \; F(\mathbf{x}) + G(\mathbf{z}) \text{ s.t. } \mathbf{K}\mathbf{x} = \mathbf{z}$$

Forward-Backward Splitting [2]

$$\begin{cases} \mathbf{y}^{(k+1)} = \mathbf{x}^{(k)} - \lambda \mathbf{K}^* \nabla G(\mathbf{K}\mathbf{x}^{(k)}) \\ \mathbf{x}^{(k+1)} = \mathrm{prox}_{\lambda F}(\mathbf{y}^{(k+1)}) \end{cases}$$

One parameter to adjust: λ

Alternating Direction Method of Multipliers (ADMM) [2,3]

$$\begin{cases} \mathbf{x}^{(k+1)} = \mathrm{prox}^{\mathbf{K}}_{\lambda F}(\mathbf{z}^{(k)} - \mathbf{y}^{(k)})) \\ \mathbf{z}^{(k+1)} = \mathrm{prox}_{\lambda G}(\mathbf{K}\mathbf{x}^{(k+1)} + \mathbf{y}^{(k)}) \\ \mathbf{y}^{(k+1)} = \mathbf{y}^{(k)} + \mathbf{K}\mathbf{x}^{(k+1)} - \mathbf{z}^{(k+1)} \end{cases}$$

Proximal Operators [2]: $\mathrm{prox}^{\mathbf{K}}_{\lambda f}\,\mathbf{z} = \arg\min_{\mathbf{x}} \lambda f(\mathbf{x}) + \frac{1}{2}\|\mathbf{K}\mathbf{x} - \mathbf{z}\|^2$

[2] P. L. Combettes and J. C. Pesquet. *Proximal splitting methods in signal processing*. Fixed-Point Algorithms for Inverse Problems in Science and Engineering, 185-212. 2011.
[3] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. *Distributed optimization and statistical learning via the alternating direction method of multipliers*. Foundations and Trends®in Machine Learning, **3**(1), 2011.

Iterative algorithms

$$\min_{\mathbf{x}\in\mathcal{X}} \; F(\mathbf{K}\mathbf{x}) + G(\mathbf{x})$$

Chambolle-Pock (CP) [4]

$$\begin{cases} \mathbf{y}^{(k+1)} & = \mathrm{prox}_{\sigma F^\star}(\mathbf{y}^{(k)} + \sigma \mathbf{K}\bar{\mathbf{x}}^{(k)}) \\ \mathbf{x}^{(k+1)} & = \mathrm{prox}_{\tau G}(\mathbf{x}^{(k)} - \tau \mathbf{K}^* \mathbf{y}^{(k+1)}) \\ \bar{\mathbf{x}}^{(k+1)} & = \mathbf{x}^{(k+1)} + \vartheta(\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}) \end{cases}$$

Primal-Dual Hybrid Gradient (PDHG) [5]

$$\begin{cases} \mathbf{x}^{(k+1)} & = \mathrm{prox}_{\tau G}(\mathbf{x}^{(k)} - \tau \mathbf{K}^* \mathbf{y}^{(k)}) \\ \bar{\mathbf{x}}^{(k+1)} & = 2\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)} \\ \mathbf{y}^{(k+1)} & = \mathrm{prox}_{\sigma F^\star}(\mathbf{y}^{(k)} + \sigma \mathbf{K}\bar{\mathbf{x}}^{(k+1)}) \end{cases}$$

Two parameters to adjust: σ, τ

[4] A. Chambolle and T. Pock. *A first-order primal-dual algorithm for convex problems with applications to imaging*. Journal of Mathematical Imaging and Vision **40**(1), 120–145. 2011
[5] Zhu, M., and Chan, T.F., An Efficient Primal-Dual Hybrid Gradient Algorithm for Total Variation Image Restoration, UCLA CAM Report [08-34], May 2008.

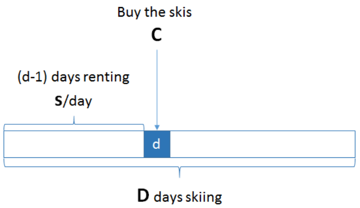
Outline

1. General optimization problems
2. Iterative algorithms
3. Step-size selection using ski-rental problem
4. Step-size selection using residuals balancing
5. Numerical examples

Ski-rental problem

Principle

Online strategy for choosing between paying a **repeated cost** versus paying a **one-time cost** to perform a task.



Optimal day d to buy the skis in order to minimize the cost:

$$d = \min\left\{\frac{C}{s}, D\right\}$$

D is unknown

Break-even strategy

$$sd = C \quad \Rightarrow \quad d = \frac{C}{s}$$

[6] A. Aghazadeh, A. Ayremlou, D.D. Calderón, T. Goldstein, R. Patel, D. Vats, R.G. Baraniuk. Adaptive step-size selection for optimization via the ski rental problem. ICASSP 2013.

Step-size selection using ski-rental problem

Goal

Optimal schedule for updating the step-size.

- Renting: d iterations using the original step-size
- Buying: $C + (1 - F)d \Leftrightarrow$ updating cost plus the amount of iterations after the step-size update, for some $F \in (0, 1)$

Break-even strategy

$$d = C + (1 - F)d \quad \Rightarrow \quad d = \frac{C}{F} \text{ iterations}$$

[6] A. Aghazadeh, A. Ayremlou, D.D. Calderón, T. Goldstein, R. Patel, D. Vats, R.G. Baraniuk. Adaptive step-size selection for optimization via the ski rental problem. ICASSP 2013.

Step-size selection using ski-rental problem

Initialize $F = 0.1$; $t > 0$; $\gamma = 2$; $C = 3$; $k = 0$; $r = 2\varepsilon$

1: $r_0 = r$; $k = k + 1$; $B = \frac{C}{F}$

2: Compute algorithm with current stepsize (t)

$$\{\mathbf{x}, r\} = \mathrm{Algorithm}(\mathbf{x}, t)$$
$$I = \frac{1}{\log_{10}(r/r_0)}$$

3: **if** $k = B$ **then** Update the stepsize

$$t' = [t/\gamma \quad t \quad t\gamma]$$
$$\{\mathbf{x}_1, r_1\} = \mathrm{Algorithm}(\mathbf{x}, t'_1)$$
$$\{\mathbf{x}_2, r_2\} = \mathrm{Algorithm}(\mathbf{x}, t'_2)$$
$$\{\mathbf{x}_3, r_3\} = \mathrm{Algorithm}(\mathbf{x}, t'_3)$$
$$i = \arg\min_j r_j;\ I_0 = I;\ I = \frac{1}{\log_{10}(r_i/r)};\ F = 1 - \frac{I}{I_0};\ k = 0;$$
$$\mathbf{x} = \mathbf{x}_i;\ r = r_i;\ t = t'_i$$

4: Stop is the residual is small enough

if $(r \geq \varepsilon)$ **then** break

Outline

1 General optimization problems

2 Iterative algorithms

3 Step-size selection using ski-rental problem

4 Step-size selection using residuals balancing

5 Numerical examples

Primal-Dual Residuals

Primal problem

$$\min_{\mathbf{x}\in\mathcal{X}} \; F(\mathbf{Kx}) + G(\mathbf{x})$$

Dual problem

$$\max_{\mathbf{y}\in\mathcal{Y}} -F^\star(\mathbf{y}) - G^\star(-\mathbf{K}^*\mathbf{y})$$

Saddle point problem

$$\min_{\mathbf{x}\in\mathcal{X}} \max_{\mathbf{y}\in\mathcal{Y}} \langle \mathbf{Kx}, \mathbf{y}\rangle + G(\mathbf{x}) - F^\star(\mathbf{y})$$

Residuals

Primal residual

$$P(\mathbf{x},\mathbf{y}) := \partial G(\mathbf{x}) + \mathbf{K}^*\mathbf{y}$$

Dual residual

$$D(\mathbf{x},\mathbf{y}) := \partial F^\star(\mathbf{y}) - \mathbf{Kx}$$

P **and** D **must tend to zero!!**

[3] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. *Distributed optimization and statistical learning via the alternating direction method of multipliers*. Foundations and Trends® in Machine Learning, **3**(1), 2011.
[7] T. Goldstein, Esser E., and R. Baraniuk, Adaptive primal-dual hybrid gradient methods for saddle-point problems, preprint, arXiv:1305.0546v1 (2013).

Optimality conditions CP algorithm

- CP algorithm

$$\begin{cases} \mathbf{y}^{(k+1)} & = \text{prox}_{\sigma F^\star}(\mathbf{y}^{(k)} + \sigma \mathbf{K}\bar{\mathbf{x}}^{(k)}) \\ \mathbf{x}^{(k+1)} & = \text{prox}_{\tau G}(\mathbf{x}^{(k)} - \tau \mathbf{K}^* \mathbf{y}^{(k+1)}) \\ \bar{\mathbf{x}}^{(k+1)} & = \mathbf{x}^{(k+1)} + \vartheta(\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}) \end{cases}$$

$$\Downarrow$$

$$\begin{cases} \mathbf{y}^{(k+1)} & = \arg\min_{\mathbf{z}} \sigma F^\star(\mathbf{z}) + \frac{1}{2}\|\mathbf{z} - \mathbf{y}^{(k)} + \sigma \mathbf{K}(\mathbf{x}^{(k)}(1+\vartheta) - \vartheta \mathbf{x}^{(k-1)})\|^2 \\ \mathbf{x}^{(k+1)} & = \arg\min_{\mathbf{z}} \tau G(\mathbf{z}) + \frac{1}{2}\|\mathbf{z} - \mathbf{x}^{(k)} - \tau \mathbf{K}^* \mathbf{y}^{(k+1)}\|^2 \end{cases}$$

- Optimality Conditions

$$0 \in \sigma \partial F^\star(\mathbf{y}^{(k+1)}) + \mathbf{y}^{(k+1)} - \mathbf{y}^{(k)} - \sigma \mathbf{K}(\mathbf{x}^{(k)}(1+\vartheta) - \vartheta \mathbf{x}^{(k-1)})$$

$$0 \in \tau \partial G(\mathbf{x}^{(k+1)}) + \mathbf{x}^{(k+1)} - \mathbf{x}^{(k)} + \tau \mathbf{K}^* \mathbf{y}^{(k+1)}$$

Optimality conditions CP algorithm

- Optimality Conditions

$$0 \in \sigma\partial F^\star(\mathbf{y}^{(k+1)}) + \mathbf{y}^{(k+1)} - \mathbf{y}^{(k)} - \sigma\mathbf{K}(\mathbf{x}^{(k)}(1+\vartheta) - \vartheta\mathbf{x}^{(k-1)})$$
$$0 \in \tau\partial G(\mathbf{x}^{(k+1)}) + \mathbf{x}^{(k+1)} - \mathbf{x}^{(k)} + \tau\mathbf{K}^*\mathbf{y}^{(k+1)}$$
$$\Downarrow$$
$$\frac{1}{\sigma}(\mathbf{y}^{(k)} - \mathbf{y}^{(k+1)}) + \mathbf{K}((1+\vartheta)\mathbf{x}^{(k)} - \vartheta\mathbf{x}^{(k-1)} - \mathbf{x}^{(k+1)}) \in \partial F^\star(\mathbf{y}^{(k+1)}) - \mathbf{K}\mathbf{x}^{(k+1)}$$
$$\frac{1}{\tau}(\mathbf{x}^{(k)} - \mathbf{x}^{(k+1)}) \in \partial G(\mathbf{x}^{(k+1)}) + \mathbf{K}^*\mathbf{y}^{(k+1)}$$

- Residuals

$$D(\mathbf{x},\mathbf{y}) = \partial F^\star(\mathbf{y}) - \mathbf{K}\mathbf{x}$$
$$P(\mathbf{x},\mathbf{y}) = \partial G(\mathbf{x}) + \mathbf{K}^*\mathbf{y}$$

$$\mathbf{P}^{(k+1)} = \frac{1}{\tau}(\mathbf{x}^{(k)} - \mathbf{x}^{(k+1)})$$
$$\mathbf{D}^{(k+1)} = \frac{1}{\sigma}(\mathbf{y}^{(k)} - \mathbf{y}^{(k+1)}) + \mathbf{K}\big((1+\vartheta)\mathbf{x}^{(k)} - \vartheta\mathbf{x}^{(k-1)} - \mathbf{x}^{(k+1)}\big)$$

Step-size selection CP algorithm

Tradeoff between the primal and dual residuals

$\tau \to \mathbf{P}$ then τ big $\Rightarrow$ Fast primal minimization but slow dual minimization
$\sigma \to \mathbf{D}$ then σ big $\Rightarrow$ Fast dual minimization but slow primal minimization
Idea: choose stepsizes so that the larger $\mathbf{P}^{(k)}$ and $\mathbf{D}^{(k)}$ are as small as possible.

Residual balancing

Assuming $\mathbf{P}^{(k)}/\mathbf{D}^{(k)}$ decrease/increase monotonically with τ, then $\max\{\mathbf{P}^{(k)}, \mathbf{D}^{(k)}\}$ is minimized when

$$\|\mathbf{P}^{(k)}\|_1 = c\|\mathbf{D}^{(k)}\|_1$$

τ and σ are chosen to "balance" the primal and dual residuals.

The scaling parameter c depends on the dynamics (dyn) of the problem input.
Suggestion [2]: $c = 255/\text{dyn}$.

[3] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. *Distributed optimization and statistical learning via the alternating direction method of multipliers*. Foundations and Trends® in Machine Learning, **3**(1), 2011.
[7] T. Goldstein, Esser E., and R. Baraniuk, Adaptive primal-dual hybrid gradient methods for saddle-point problems, preprint, arXiv:1305.0546v1 (2013).

Step-size selection CP algorithm

Initialize $\tau^{(1)}, \sigma^{(1)} = \frac{1}{\|K\|}$; $\alpha^{(1)} = 0.5$; $\eta = 0.95$; $\Delta = 1.5$

1: Compute CP iterations using $\tau^{(k)}$, $\sigma^{(k)}$.

2: Compute the ℓ_1-norm of the Primal and Dual Residuals

$$p = \|\mathbf{P}^{(k)}\|_1 \qquad d = \|\mathbf{D}^{(k)}\|_1$$

3: Update parameters:

if $(p > c\Delta d)$ **then**
$\tau^{(k+1)} = \tau^{(k)}(1-\alpha^{(k)})$; $\sigma^{(k+1)} = \sigma^{(k)}/(1-\alpha^{(k)})$; $\alpha^{(k+1)} = \alpha^{(k)}\eta$
else if $(p < c\Delta d)$ **then**
$\tau^{(k+1)} = \tau^{(k)}/(1-\alpha^{(k)})$; $\sigma^{(k+1)} = \sigma^{(k)}(1-\alpha^{(k)})$; $\alpha^{(k+1)} = \alpha^{(k)}\eta$
else
$\tau^{(k+1)} = \tau^{(k)}$; $\sigma^{(k+1)} = \sigma^{(k)}$; $\alpha^{(k+1)} = \alpha^{(k)}$

4: Stop if the algorithm is stable:
if $(p \leq \mathtt{Th})\&\&(d \leq \mathtt{Th})$ **then** break

Outline

1 General optimization problems

2 Iterative algorithms

3 Step-size selection using ski-rental problem

4 Step-size selection using residuals balancing

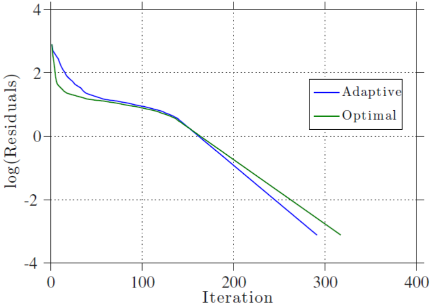
5 Numerical examples

Sparse regression problem using FBS - Ski-rental problem

$$\mathbf{y} = \mathbf{K}\mathbf{x} + \mathbf{n}$$

- $\mathbf{K} \in \mathbb{R}^{100 \times 1000}$ random Gaussian matrix
- $\mathbf{x} \in \mathbb{R}^{1000}$ 30-sparse random vector
- $\mathbf{n} \in \mathbb{R}^{100}$, $n_i \sim \mathcal{N}(0, 10^{-6})$

$$\min_{\mathbf{x}} \mu \|\mathbf{x}\|_1 + \frac{1}{2}\|\mathbf{K}\mathbf{x} - \mathbf{y}\|_2^2$$



$\mu = 10;\ t_{\text{opt}} = 0.0015$

[6] A. Aghazadeh, A. Ayremlou, D.D. Calderón, T. Goldstein, R. Patel, D. Vats, R.G. Baraniuk. Adaptive step-size selection for optimization via the ski rental problem. ICASSP 2013.

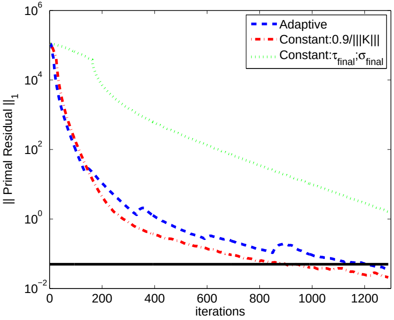
TV-Denoising using CP - Residual balancing

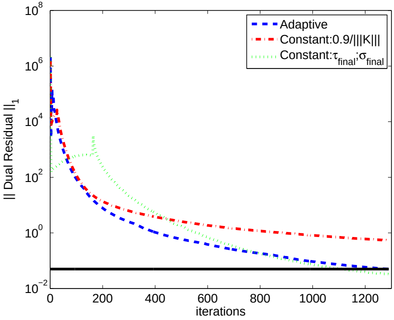
$$\mathbf{y} = \mathbf{x} + \mathbf{n}$$

$$\min_{\mathbf{x}} \|\mathbf{x}\|_{TV} \text{ s.t. } \frac{1}{2}\|\mathbf{x} - \mathbf{y}\|_2^2 \leq \varepsilon$$

- $\mathbf{x} \in \mathbb{R}^{256^2}$: cameraman $\in [0, 255]$
- $\mathbf{n} \in \mathbb{R}^{256^2}$, $n_i \sim \mathcal{N}(0, 100)$

- Adaptive: $[\sigma, \tau]^{(1)} = \frac{0.9}{|||\mathbf{K}|||} \rightarrow [\sigma, \tau]_{\text{final}}$
- Constant: $\sigma, \tau = \frac{0.9}{|||\mathbf{K}|||}$
- Constant: $\sigma_{\text{final}} = 2.09$; $\tau_{\text{final}} = 0.049$





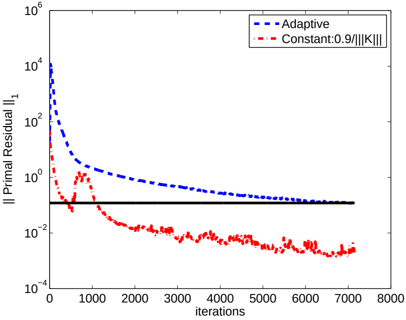
4 times less iterations!!

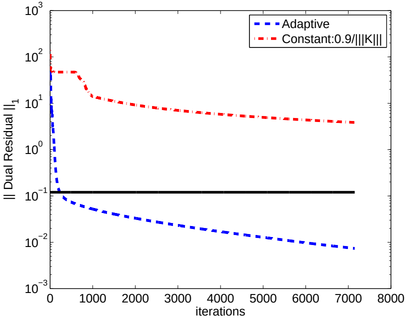
$$\mathbf{y} = \mathbf{\Phi x} + \mathbf{n}$$

- $\mathbf{\Phi} = \mathbf{DF} \in \mathbb{C}^{(360*397)\times 256^2}$
- $\mathbf{x} \in \mathbb{R}^{256^2}$: bundle of fibers $\in [0, 0.012]$
- $\mathbf{n} \in \mathbb{R}^{256^2}$, $n_i \sim \mathcal{N}(0, 10^{-8})$

$$\min_{\mathbf{x}} \|\mathbf{x}\|_{TV} \text{ s.t. } \begin{cases} \|\mathbf{\Phi x} - \mathbf{y}\|_2 \leq \varepsilon \\ \mathbf{x} \succeq 0 \\ \mathbf{x}_\Omega = 0 \end{cases}$$

- Adaptive: $[\sigma, \tau]^{(1)} = \frac{0.9}{|||\mathbf{K}|||} \to [\sigma, \tau]_{\text{final}}$
- Constant: $\sigma, \tau = \frac{0.9}{|||\mathbf{K}|||}$





7 times less iterations!!

References

1 A. González, L. Jacques, C. De Vleeschouwer, and P. Antoine, Compressive optical deflectometric tomography: A constrained total-variation minimization approach. To appear in Inverse Problems and Imaging Journal, 2014, preprint, arXiv:1209.0654.

2 P. L. Combettes and J. C. Pesquet. *Proximal splitting methods in signal processing*. Fixed-Point Algorithms for Inverse Problems in Science and Engineering, 185-212. 2011.

3 S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein. *Distributed optimization and statistical learning via the alternating direction method of multipliers*. Foundations and Trends® in Machine Learning, **3**(1), 2011.

4 A. Chambolle and T. Pock. *A first-order primal-dual algorithm for convex problems with applications to imaging*. Journal of Mathematical Imaging and Vision **40**(1), 120–145. 2011.

5 Zhu, M., and Chan, T.F., An Efficient Primal-Dual Hybrid Gradient Algorithm for Total Variation Image Restoration, UCLA CAM Report [08-34], May 2008.

6 A. Aghazadeh, A. Ayremlou, D.D. Calderón, T. Goldstein, R. Patel, D. Vats, R.G. Baraniuk. Adaptive step-size selection for optimization via the ski rental problem. ICASSP 2013.

7 T. Goldstein, Esser E., and R. Baraniuk, Adaptive primal-dual hybrid gradient methods for saddle-point problems, preprint, arXiv:1305.0546v1 (2013).

Thank you!!