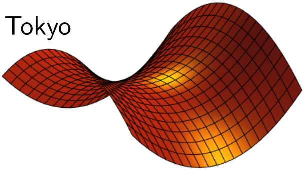
New Horizon of Mathematical Sciences
RIKEN iTHES - Tohoku AIMR - Tokyo IIS Joint Symposium
April 28 2016

AUGMENTED LAGRANGIAN METHODS FOR CONVEX OPTIMIZATION

T. Takeuchi

Aihara Lab.
Laboratories for Mathematics, Lifesciences, and Informatics

Institute of Industrial Science, the University of Tokyo



Optimization Problems: Types and Algorithms[1]

- Types of Optimization
  - Linear Programming
  - Quadratic Programming
  - Semidefinite Programming
  - MPEC
  - Nonlinear Least Squares
  - Nonlinear Equations
  - Nondifferentiable Optimization
  - Global Optimization
  - Derivative-Free Optimization
  - Network Optimization

- Algorithms
  - Line Search
  - Trust-Region
  - (Nonlinear) Conjugate Gradient
  - (Quasi-)Newton
  - Truncated Newton
  - Gauss Newton
  - Levenberg-Marquardt
  - Homotopy
  - Gradient Projection
  - Simplex Method
  - Interior Point Methods
  - Active Set Methods
  - Primal-Dual Active Set Methods
  - Augmented Lagrangian Methods
  - Reduced Gradient
  - Sequential Quadratic Programming

[1] neos Optimization Guide: *http://www.neos-guide.org*

Problem and Objective

- Problem: A class of nonsmooth convex optimization

$$\min_x f(x) + \phi(Ex) \tag{1}$$

  - X, H: real Hilbert spaces
  - $f\colon X \to \mathbb{R}$: differentiable, $\nabla^2 f(x) \succeq mI$
  - $\phi\colon H \to \mathbb{R}$: convex, nonsmooth(nondifferentiable)
  - $E\colon X \to H$: linear

⋆ Optimality system

$$0 \in \nabla_x f(x) + E^\top \partial\phi(Ex) \tag{2}$$

- Objective: To develop Newton (-like) methods
  - Applicable to a wide range of applications
  - Fast and accurate
  - Easy to implement
- Tool: **Augmented Lagrangian in the sense of Fortin.**

Motivating Examples (1/4)

- Optimal control problem

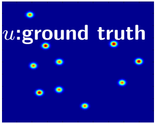
$$\begin{aligned} \min \quad & \frac{1}{2}\int_0^T x(t)^\top Q x(t) + u(t)^\top R u(t)\, dt \\ \text{s.t.} \quad & \dot{x}(t) = Ax(t) + Bu(t), \quad x(0) = x_0 \\ & \|u(t)\| \le 1. \end{aligned}$$

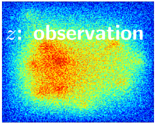
- Inverse source problem

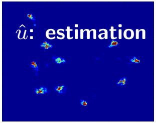
$$(\hat{y}, \hat{u}) = \arg \min_{(y,u)\in H_0^1(\Omega)\times L^2(\Omega)} \quad (1/2)\|y - z\|_{L^2}^2 + (\alpha/2)\|u\|_{L^2}^2 + \beta\|u\|_{L^1}$$

$$\text{s.t.} \quad -\Delta y = u \quad \text{on} \quad \Omega,$$

$$y = 0 \quad \text{on} \quad \partial\Omega$$







Motivating Examples (2)

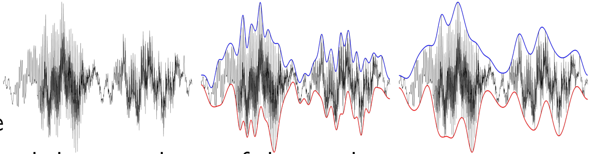
- lasso (least absolute shrinkage and selection operator) [R.Tibshinari, 96]

$$\min_{\beta \in \mathbb{R}^p} \|X\beta - y\|_2^2 \text{ subject to } \|\beta\|_1 \leq t. \quad \left(\min_{\beta \in \mathbb{R}^p} \|X\beta - y\|_2^2 + \lambda \|\beta\|_1 \right)$$

  - $X = [\boldsymbol{x}^1, \ldots, \boldsymbol{x}^N]^\top$
  - $\boldsymbol{x}^i = (x_{i1}, \ldots, x_{ip})^\top$: the predictor variables, $i = 1, 2, \ldots, N$
  - y_i: the responses, $i = 1, 2, \ldots, N$

- envelope construction

$$\color{blue}{\hat{x}} \color{black} = \arg\min_{x \in \mathbb{R}^n} \frac{1}{2}\|x - s\|_2^2 + \alpha \|(D^\top D)^2 x\|_1 \text{ subject to } s \leq x.$$



  - s: noisy signal
  - D: finite difference
  - $\|(D^\top D)^2 x\|_1$: control the smoothness of the envelope x.

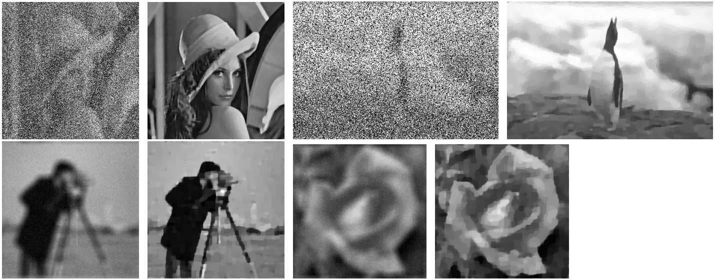
Motivating Examples (3/4)

- total variation image reconstruction

$$\min \frac{1}{p}\|Kx-y\|_{L^p} + \alpha|x|_{TV} + \beta/2\|x\|_{H^1}^2$$

for $p = 1$ or $p = 2$. K: blurring kernel.

$$|x|_{TV} = \sum_{i,j}\sqrt{([D_1 x]_{i,j})^2 + ([D_2 x]_{i,j})^2}$$



Motivating Examples (4/4)

- Quadratic programming

$$\min_x \frac{1}{2}(x, Ax) - (a, x) \quad \text{s.t.} \quad Ex = b, \quad Gx \geq g.$$

- Group LASSO

$$\min_\beta \frac{1}{2}\|X\beta - y\|^2 + \alpha\|\beta\|_G,$$

e.g. $G = \{\{1,2\},\{3,4,5\}\}$, $\|\beta\|_G = \sqrt{\beta_1^2 + \beta_2^2} + \sqrt{\beta_3^2 + \beta_4^2 + \beta_5^2}$.

- SVM
- $\cdots$

outline

1. Augmented Lagrangian

2. Newton Methods

3. Numerical Test

Augmented Lagrangian

Method of multipliers for equality constraint

- equality constrained optimization problem

$$(P) \qquad \min_x \; f(x) \text{ subject to } g(x) = 0$$

where $f\colon \mathbb{R}^n \to \mathbb{R}$ and $g\colon \mathbb{R}^n \to \mathbb{R}^m$ are differentiable.

- augmented Lagrangian (Hestenes, Powell 69), $c > 0$

$$\begin{aligned} L_c(x,\lambda) &= f(x) + \lambda^\top g(x) + (c/2)\|g(x)\|_{\mathbb{R}^m}^2 \\ &= f(x) + (c/2)\|g(x) + \lambda/c\|_{\mathbb{R}^m}^2 - 1/(2c)\|\lambda\|_{\mathbb{R}^m}^2 . \end{aligned}$$

- method of multiplier[1]

$$\lambda_{k+1} = \lambda_k + c[\nabla_\lambda d_c](\lambda_k).$$

⋆ method of multiplier is composed of two steps:

$$\begin{aligned} x^k &= \arg\min_x L_c(x, \lambda^k) \\ \lambda_{k+1} &= \lambda_k + c\nabla_\lambda L_c(x^k, \lambda^k) = \lambda_k + cg(x^k) \end{aligned}$$

[1]**def**.: $d_c(\lambda) := \min_x L_c(x,\lambda)$. **fact**: $\nabla d_c(\lambda) = [\nabla_\lambda L_c](x(\lambda),\lambda) = g(x(\lambda))$

A brief history of augmented Lagrangian methods (1/2)

■ equality constrained optimization

$$\min_{x\in\mathbb{R}^n} \quad f(x)$$
$$\text{s.t.} \quad h(x) = 0$$

- 1969 Hestenes[7], Powell[14]

■ inequality constrained optimization

$$\min_{x\in\mathbb{R}^n} \quad f(x)$$
$$\text{s.t.} \quad g(x) \leq 0$$

- 1973 Rockafellar[15, 16]

■ equality and inequality constrained optimization

$$\min_{x\in\mathbb{R}^n} \quad f(x)$$
$$\text{s.t.} \quad h(x) = 0$$
$$\quad g(x) \leq 0$$

- 1982 Berstekas[2]

■ nonsmooth convex optimization

$$\min_{x\in\mathbb{R}^n} \quad f(x) + \phi(Ex)$$

- 1975 Glowinski-Marroco[6]
- 1975 Fortin[4]
- 1976 Gabay-Mercier[5]
- 2000 Ito-Kunisch[9]
- 2009 Tomioka-Sugiyama[18]

A brief history of augmented Lagrangian methods (2/2)

- Augmented Lagrangian Methods (Method of Multipliers)

year	authors	problem	algorithm
1969	Hestenes[7], Powell[14]	equality constraint	
1973	Rockafellar[15, 16]	inequality constraint	
1975	Glowinski-Marroco[6]	nonsmooth convex (FEM)	ALG1
1975	Fortin[4]	nonsmooth convex (FEM)	
1976	Gabay-Mercier[5]	nonsmooth convex (FEM)	ADMM[1]
1982	Berstekas[2]	equality and inequality constraints	
2000	Ito-Kunisch[9]	nonsmooth convex (PDE Opt.)	
2009	Tomioka-Sugiyama[18]	nonsmooth convex (Machine Learning)	DAL[2]
2013	Patrinos-Bemporad [12]	convex composite	Proximal Newton
2014	Patrinos-Stella -Bemporad[13]	(Optimal Control)	FBN[3]

[1] Alternating Direction Method of Multipliers
[2] Dual-augmented Lagrangian method
[3] Forward-Backward-Newton

Augmented Lagrangian Methods [Glowinski+,75]

- nonsmooth convex optimization problem

$$\min_{x,v} \; f(x) + \phi(v), \quad \text{subject to} \quad v = Ex.$$

- augmented Lagrangian

$$L_c(x, v, \lambda) = f(x) + \phi(v) + (\lambda, Ex - v) + (c/2)\|Ex - v\|^2_{\mathbb{R}^m}$$

- dual function

$$d_c(\lambda) := \min_{x,v} L_c(x, v, \lambda)$$

- method of multipliers (ALG1)

$$\lambda_{k+1} = \lambda_k + c\nabla_\lambda d_c(\lambda_k).$$

⋆ method of multiplier is composed of two steps:

step 1	(x_{k+1}, v_{k+1})	$=$	$\arg\min_{x,v} L_c(x, v, \lambda_k)$
step 2	λ_{k+1}	$=$	$\lambda_k + c(Ex_{k+1} - v_{k+1})$

Fact (gradient): $\nabla_\lambda d_c(\lambda) = [\nabla_\lambda L_c](x(\lambda), v(\lambda), \lambda)$

Augmented Lagrangian Methods [Gabay+,76]

- nonsmooth convex optimization problem

$$\min_{x,v} \; f(x) + \phi(v), \quad \text{subject to} \quad v = Ex.$$

- augmented Lagrangian

$$L_c(x, v, \lambda) = f(x) + \phi(v) + (\lambda, Ex - v) + (c/2)\|Ex - v\|^2_{\mathbb{R}^m}$$

- ADMM, alternating direction method of multiplier

$$\begin{array}{lrcl} \text{step 1} & x_{k+1} & = & \arg\min_x L_c(x, v_k, \lambda_k) \\ \text{step 2} & v_{k+1} & = & \arg\min_v L_c(x_{k+1}, v, \lambda_k) = \text{prox}_{\phi/c}(Ex_{k+1} + \lambda_k/c) \\ \text{step 3} & \lambda_{k+1} & = & \lambda_k + c(Ex_{k+1} - v_{k+1}) \end{array}$$

Assumption: $\text{prox}_{\phi/c}(z) := \arg\min_v(\phi(v) + c/2\|v - z\|^2)$ has a closed-form representation (explicitly computable).

Augmented Lagrangian Methods [Fortin,75]

- nonsmooth convex optimization problem

$$\min_x \; f(x) + \phi(Ex)$$

- (Fortin's) augmented Lagrangian

$$L_c(x,\lambda) = \min_v L_c(x,v,\lambda) = f(x) + \phi_c(Ex + \lambda/c) - 1/(2c)\|\lambda\|^2_{\mathbb{R}^m}$$

where $\phi_c(z)$ is Moreau's envelope

- method of multiplier

$$\lambda_{k+1} = \lambda_k + c\nabla_\lambda d_c(\lambda_k)$$

def.(dual function): $d_c(\lambda) := \min_x L_c(x,\lambda)$.

⋆ method of multiplier is composed of two steps:

$$x_k = \arg\min_x L_c(x,\lambda_k)$$

$$\lambda_{k+1} = \lambda_k + c[Ex^k - \text{prox}_{\phi/c}(Ex^k + \lambda^k/c)]$$

fact (gradient): $\nabla_\lambda d_c(\lambda) = [\nabla_\lambda L_c](x(\lambda),\lambda)$

Assumption: $\text{prox}_{\phi/c}(z) := \arg\min_v(\phi(v) + c/2\|v-z\|^2)$ has a closed-form representation (explicitly computable).

Comparison

- augmented Lagrangian for ALG1[Glowinski+,75], ADMM[Gabay+,76]

$$L_c(x, v, \lambda) = f(x) + \phi(v) + (\lambda, Ex - v) + (c/2)\|Ex - v\|^2_{\mathbb{R}^m}$$

- (Fortin's) augmented Lagrangian for ALM[Fortin,75]

$$\begin{aligned} L_c(x, \lambda) &= \min_v L_c(x, v, \lambda) \\ &= f(x) + \phi_c(Ex + \lambda/c) - 1/(2c)\|\lambda\|^2_{\mathbb{R}^m} \end{aligned}$$

- Fortin's Augmented Lagrangian method has been ignored in the subsequent literature.
- The method was re-discovered by [Ito-Kunisch,2000] and by [Tomioka-Sugiyama,2009].

Moreau's envelope, proximal operator (1/3)

- Moreau's envelope (Moreau-Yosida regularization) of a closed proper convex function ϕ

$$\phi_c(z) := \min_v(\phi(v) + c/2\|v - z\|^2)$$

- The proximal operator $\mathsf{prox}_\phi(z)$ is defined by

$$\mathsf{prox}_\phi(z) := \arg\min_v(\phi(v) + 1/2\|v - z\|^2)$$

$$\begin{aligned}\mathsf{prox}_{\phi/c}(z) &= \arg\min_v(\phi(v) + c/2\|v - z\|^2)\\ &= \arg\min_v(\phi(v)/c + 1/2\|v - z\|^2)\end{aligned}$$

- Moreau's envelope is expressed in terms of the proximal operator

$$\phi_c(z) = \phi(\mathsf{prox}_{\phi/c}(z)) + c/2\|\mathsf{prox}_{\phi/c}(z) - z\|^2$$

Moreau's envelope, proximal operator (2/3)

•ℓ_1 norm

$$\phi(x) = \|x\|_1,\ x \in \mathbb{R}^n$$

$$\text{prox}_\phi(x)_i = \max(0, |x_i| - 1)\text{sign}(x_i),$$

$$G \in \partial_B(\text{prox}_\phi)(x)$$

$$G_{ii} = \begin{cases} 1 & |x_i| > 1, \\ 0 & |x_i| < 1, \\ \in \{0,1\} & |x_i| = 1 \end{cases}$$

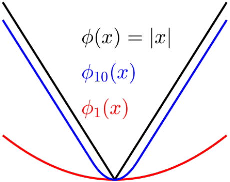
•Euclidean norm

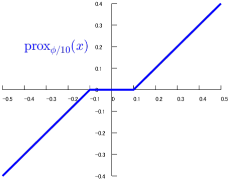
$$\phi(x) = \|x\|,\ x \in \mathbb{R}^n$$

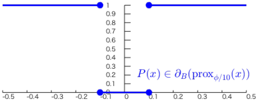
$$\text{prox}_\phi(x) = \max(0, \|x\| - 1)\frac{x}{\|x\|},$$

$$\partial_B(\text{prox}_\phi)(x) =$$

$$\begin{cases} I - \frac{1}{\|x\|}(I - \frac{xx^\top}{\|x\|^2}) & \|x\| > 1, \\ 0 & \|x\| < 1, \\ \{0, I - \frac{1}{\|x\|}(I - \frac{xx^\top}{\|x\|^2})\} & \|x\| = 1 \end{cases}$$







Ref. Fixed-Point Algorithms for Inverse Problems in Science and Engineering Springer Optimization and Its Applications, Chapter 10.

Moreau's envelope, proximal operator (3/3)

- Properties[1]
  - $\phi_c(z) \le \phi(z)$, $\lim_{c\to\infty} \phi_c(z) = \phi(z)$
  - $\phi_c(z) + (\phi^*)_{1/c}(cz) = (c/2)\|z\|^2$

def.(convex conjugate):
$\phi^*(x^*) := \sup_x(\langle x^*, x\rangle - \phi(x))$.

  - $\nabla_z \phi_c(z) = c(z - \mathsf{prox}_{\phi/c}(z))$
  - $\|\mathsf{prox}_{\phi/c}(z) - \mathsf{prox}_{\phi/c}(w)\|^2 \le \langle \mathsf{prox}_{\phi/c}(z) - \mathsf{prox}_{\phi/c}(w), z - w)$
  - $\mathsf{prox}_{\phi/c}(z) + (1/c)\mathsf{prox}_{c\phi^*}(cz) = z$
  - $\mathsf{prox}_{\phi/c}$ is globally Lipschitz, almost everywhere differentiable
  - $P \in \partial_B(\mathsf{prox}_{\phi/c})(x)$ is a symmetric positive semidefinite that satisfies $\|P\| \le 1$.
  - $\partial_B(\mathsf{prox}_{c\phi^*})(x) = \{I - P \colon P \in \partial_B(\mathsf{prox}_{\phi/c})(x/c)\}$

[1] [17, Bauschke+11],[13, Partrions+14]

Differentiability, Optimality condition

- Fortin's augmented Lagrangian

$$L_c(x,\lambda) = f(x) + \phi_c(Ex + \lambda/c) - 1/(2c)\|\lambda\|^2_{\mathbb{R}^m}$$

- Fortin's augmented Lagrangian L_c is convex and continuously differentiable with respect to x, and is concave and continuously differentiable with respect to λ.

$$\nabla_x L_c(x,\lambda) = \nabla_x f(x) + cE^\top(Ex + \lambda/c - \mathrm{prox}_{\phi/c}(Ex + \lambda/c)),$$
$$\nabla_\lambda L_c(x,\lambda) = Ex - \mathrm{prox}_{\phi/c}(Ex + \lambda/c).$$

- The following conditions on a pair (x^*, λ^*) are equivalent.

(a) $\quad \nabla_x f(x^*) + E^\top\lambda^* = 0 \quad$ and $\quad \lambda^* \in \partial\phi(Ex^*)$.

(b) $\quad \nabla_x f(x^*) + E^\top\lambda^* = 0 \quad$ and $\quad Ex^* - \mathrm{prox}_{\phi/c}(Ex^* + \lambda^*/c) = 0$.

(c) $\quad \nabla_x L_c(x^*, \lambda^*) = 0 \quad$ and $\quad \nabla_\lambda L_c(x^*, \lambda^*) = 0$.

(d) $\quad L_c(x^*, \lambda) \le L_c(x^*, \lambda^*) \le L_c(x, \lambda^*), \qquad \forall x \in X,\ \forall\lambda \in H$.

- If there exists a pair (x^*, λ^*) that satisfies (d), then x^* is the minimizer of (1).

Newton Methods

Newton methods

- Problem

$$\min_x f(x) + \phi(Ex)$$

- Approaches

(1-1) Lagrange-Newton: Solve the optimality system by Newton method

$$\nabla_x L_c(x,\lambda) = 0 \quad \text{and} \quad \nabla_\lambda L_c(x,\lambda) = 0$$

(1-2) A special case ($E = I$). Solve the optimality system by Newton method

$$x = \mathrm{prox}_{\phi/c}(x - \nabla_x f(x)/c)$$

(2) Apply Newton method to Fortin's augmented Lagrangian method

step 1 Solve for x_k : $\nabla_x L_c(x_k, \lambda_k) = 0$

step 2 $\lambda_{k+1} = \lambda_k + c(Ex_k - \mathrm{prox}_{\phi/c_k}(Ex_k + \lambda_k/c)).$

⋆ For fixed λ_k

$$x_k = \arg\min_x L_c(x, \lambda_k) \quad \Longleftrightarrow \quad \nabla_x L_c(x_k, \lambda_k) = 0.$$

Lagrange-Newton method

- Lagrange-Newton method: Solve the optimality system by Newton method[1]

$$\nabla_x L_c(x,\lambda) = 0 \quad \text{and} \quad \nabla_\lambda L_c(x,\lambda) = 0$$

- Newton system:

$$\left[\begin{array}{c|c} \nabla_x^2 f(x) + cE^\top (I-G)E & ((I-G)E)^\top \\ \hline (I-G)E & -c^{-1}G \end{array}\right] \begin{bmatrix} d_x \\ d_\lambda \end{bmatrix} = -\begin{bmatrix} \nabla_x L_c(x,\lambda) \\ \nabla_\lambda L_c(x,\lambda) \end{bmatrix}$$

where $G \in \partial(\mathrm{prox}_{\phi/c})(Ex + \lambda/c)$

- or equivalently

$$\left[\begin{array}{c|c} \nabla_x^2 f(x) & E^\top \\ \hline (I-G)E & -c^{-1}G \end{array}\right] \begin{bmatrix} x^+ \\ \lambda^+ \end{bmatrix} = \begin{bmatrix} \nabla_x^2 f(x)x - \nabla_x f(x) \\ \mathrm{prox}_{\phi/c}(z) - Gz \end{bmatrix}$$

where $z = Ex + \lambda/c$. $d_x = x^+ - x$, $d_\lambda = \lambda^+ - \lambda$.

[1] When E is not surjective, the Newton system becomes inconsistent, i.e., there may be no solution to the Newton system. In this case, a proximal algorithm is employed to guarantee the solvability of the Newton system

Newton method for composite convex problem

- Problem

$$\min_x f(x) + \phi(x)$$

- Apply Newton method to the optimality condition

$$\nabla_x f(x) + \lambda = 0, \quad x - \mathrm{prox}_{\phi/c}(x + \lambda/c) = 0$$
$$\iff x - \mathrm{prox}_{\phi/c}(x - \nabla_x f(x)/c) = 0$$

- Define $F_c(x) = L_c(x, -\nabla_x f(x))$.
- Algorithm

step 1: $G \in \partial(\mathrm{prox}_{\phi/c})(x - \nabla_x f(x)/c)$.
step 2: $\left(I - G(I - c^{-1}D^2 f(x))\right) d = -(x - \mathrm{prox}_{\phi/c}(x - \nabla_x f(x)/c))$.
step 3: Armijo's rule: Find the smallest nonnegative integer j such that

$$F_c(x + 2^{-j}d) \le F_c(x) + \mu 2^{-j}\langle \nabla_x F_c(x), d\rangle \quad (0 < \mu < 1)$$
$$x^+ = x + 2^{-j}d$$

⋆ step 2: a variable metric gradient method on $F_c(x)$:

$$c(I - c^{-1}\nabla_x^2 f(x))(I - G(I - c^{-1}\nabla_x^2 f(x)))d = -\nabla_x F_c(x).$$

- The algorithm is proposed in [12, Patrinos-Bemporad].

Newton method applied to the priaml update

- Apply Newton method to solve the nonlinear equation arising from Fortin's augmented Lagrangian method

$$\text{step 1} \quad \nabla_x L_{c_k}(x, \lambda_k) = 0$$

$$\text{step 2} \quad \lambda_{k+1} = \lambda_k + c_k(Ex_k - \mathrm{prox}_{\phi/c_k}(Ex_k + \lambda_k/c)).$$

- Newton system

$$(\nabla_x^2 f(x) + cE^\top (I - G)E)d_x = -\nabla_x L_{c_k}(x, \lambda)$$

Numerical Test

Numerical Test

- ℓ^1 least square problems

$$\min_x \frac{1}{2}\|Ax - y\|^2 + \gamma\|x\|_1.$$

- The data set (A, y, x_0): L1-Testset (mat files) available at `http://wwwopt.mathematik.tu-darmstadt.de/spear/`
- Used 200 datasets. (dimension of $x : 1,024 \sim 49,152$).
- Windows surface pro 3, Intel i5 / 8GB RAM.

- Comparison:
  MOSEK (free academic license available at `https://mosek.com/`)
  MOSEK is a software package for solving large scale sparse problems.
  Algorithm: the interior-point method.
- $c = 0.99/\sigma_{\max}(A^T A)$.
- $\gamma = 0.01\|A^T y\|_\infty$.
- Compute until $\sqrt{\|\nabla_x L_c(x^k, \lambda^k)\|^2 + \|\nabla_\lambda L_c(x^k, \lambda^k)\|^2} < 10^{-12}$ is met.

Newton method for ℓ^1 regularized L.S.

- Newton system:

$$\begin{bmatrix} A^\top A & I \\ I - G & -cG \end{bmatrix} \begin{bmatrix} d_x \\ d_\lambda \end{bmatrix} = - \begin{bmatrix} A^T(Ax - y) \\ x - \mathrm{prox}_{\phi/c}(x + c\lambda) \end{bmatrix}$$

where $G = \partial(\mathrm{prox}_{\phi/c})(x + c\lambda)$:

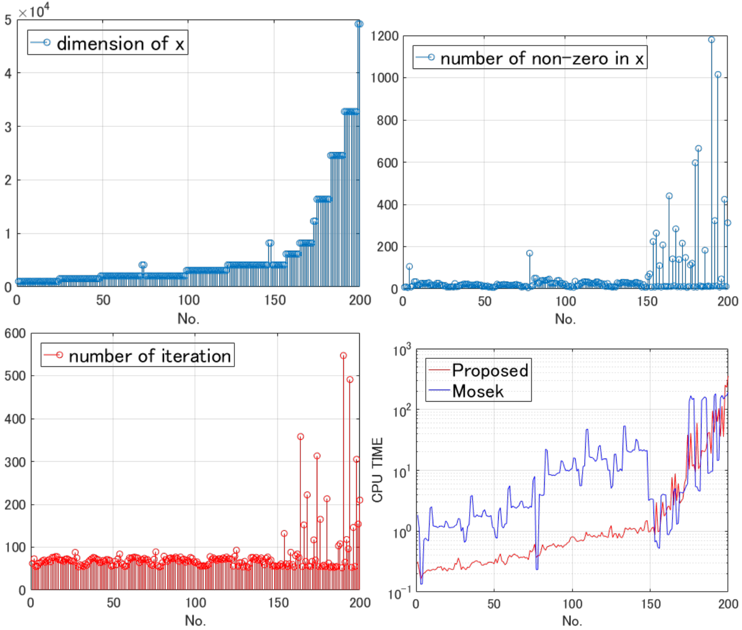
$$G_{ii} = 0 \quad \text{for } i \in \mathbf{o} := \{j \mid |x_j + c\lambda_j| \le \gamma c, j = 1, 2, \ldots, n\}$$

$$G_{ii} = 1 \quad \text{for } i \in \mathbf{o}^\mathsf{c} = \{j \mid |x_j + c\lambda_j| > \gamma c, j = 1, 2, \ldots, n\}$$

- Simply the system.

$$\begin{aligned} d_x(\mathbf{o}) &= -\, x(\mathbf{o}) \\ A(:, \mathbf{o}^\mathsf{c})^\top A(:, \mathbf{o}^\mathsf{c}) d_x(\mathbf{o}^\mathsf{c}) &= -\, [\gamma \mathsf{sign}(x(\mathbf{o}^\mathsf{c}) + c\lambda(\mathbf{o}^\mathsf{c})) \\ &\quad + A(:, \mathbf{o}^\mathsf{c})^\top (A(:, \mathbf{o}^\mathsf{c}) x(\mathbf{o}^\mathsf{c}) - y)] \\ d_\lambda &= -\, A^\top A(x + d_x) + A^T y. \end{aligned}$$

Results



×10^4
dimension of x
No.
number of non−zero in x
No.
number of iteration
No.
Proposed
Mosek
CPU TIME
No.

QUESTIONS ?
COMMENTS ?
SUGGESTIONS ?

References I

[1] Bergounioux, M., Ito, K., Kunisch, K.: Primal-dual strategy for constrained optimal control problems. SIAM J. Control Optim. **37**, 1176–1194 (1999)

[2] Bertsekas, D.: Constrained optimization and Lagrange multiplier methods. Academic Press, Inc. (1982)

[3] Facchinei, F., Pang, J.S.: Finite-Dimensional Variational Inequalities and Complementarity Problems. Vol. II. Springer-Verlag, New York (2003)

[4] Fortin, M.: Minimization of some non-differentiable functionals by the augmented Lagrangian method of Hestenes and Powell. Appl. Math. Optim. **2**, 236–250 (1975)

[5] Gaby, D., Mercier, B.: A dual algorithm for the solution of nonlinear variational problems via finite element approximation. Comp. Math. Appl. **9**, 17–40 (1976)

[6] Glowinski, R., Marroco, A.: Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité d'une classe de problèmes de dirichlet non linéaires. ESAIM: Math. Model. Numer. Anal. **9**, 41–76 (1975)

[7] Hestenes, M.R.: Multiplier and gradient methods. J. Optim. Theory Appl. **4**, 303–320 (1969)

[8] Ito, K., Kunisch, K.: An active set strategy based on the augmented Lagrangian formulation for image restoration. ESAIM: Math. Model. Numer. Anal. **33**, 1–21 (1999)

[9] Ito, K., Kunisch, K.: Augmented Lagrangian methods for nonsmooth, convex optimization in Hilbert spaces. Nonlin. Anal. Ser. A Theory Methods **41**, 591–616 (2000)

References II

[10] Ito, K., Kunisch, K.: Lagrange Multiplier Approach to Variational Problems and Applications. SIAM, Philadelphia, PA (2008)

[11] Jin, B., Takeuchi, T.: Lagrange optimality system for a class of nonsmooth convex optimization. Optimization, **65**, 1151-1166 (2016)

[12] Patrinos, P., Bemporad, A.: Proximal Newton methods for convex composite optimization. 52nd IEEE Conference on Decision and Control, 2358-2363 (2013)

[13] Patrinos, P., Stella, L., Bemporad, A.: Forward-backward truncated Newton methods for convex composite optimization. preprint, arXiv:1402.6655v2 (2014)

[14] Powell, M.J.D.: A method for nonlinear constraints in minimization problems. In: Optimization (Sympos., Univ. Keele, Keele, 1968), pp. 283–298. Academic Press, London (1969)

[15] Rockafellar, R.: A dual approach to solving nonlinear programming problems by unconstrained optimization. Math. Programming **5**, 354–373 (1973)

[16] Rockafellar, R.: The multiplier method of Hestenes and Powell applied to convex programming. J. Optim. Theory Appl. **12**, 555–562 (1973)

[17] Bauschke, H.H., Combettes, P.L.: Convex Analysis and Monotone Operator Theory in Hilbert Spaces. Springer, New York (2011)

[18] Tomioka, R., Sugiyama, M.: Dual-Augmented Lagrangian Method for Efficient Sparse Reconstruction. IEEE Signal Processing Letters. **16**, 1067–1070 (2009)