The proximal augmented Lagrangian method for non-smooth composite optimization

Mihailo Jovanović

ee.usc.edu/mihailo

Neil Dhingra

Sei Zhen Khong

NREL Optimization and Control for Autonomous Systems Workshop

Structure via regularization

$$\underset{x}{\text{minimize}} \quad \underbrace{f(x)}_{\text{performance}} \quad + \quad \underbrace{g(\mathcal{T}(x))}_{\text{structure}}$$

$$f(x) \downarrow \textbf{performance} \qquad g(\mathcal{T}(x)) \downarrow \textbf{structure}$$

- f – potentially nonconvex; Lipschitz cts gradient
- g – convex; non-differentiable

$\mathcal{T}$ – bounded linear operator
(imposes structure in desired coordinates)

Common regularizers

Examples

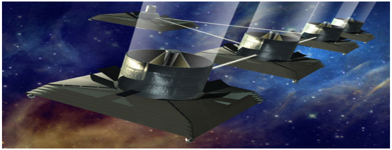
- $g(x) = I_{\mathcal{C}}(x)$ — convex constraints
- $g(x) = \|x\|_1 = \sum |x_i|$ — sparse x
- $g(x) = \|x\|_*$ — low rank x

Control applications

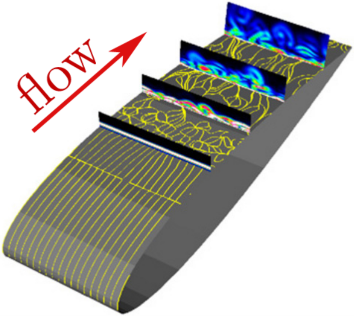
- distributed control – sparse feedback gain matrix
- sensor selection – column-sparse Kalman gain
- low-complexity modeling – low rank covariance

Motivating applications

networks of dynamical systems



fluid flows



- Challenges
  - ⋆ control-oriented modeling
  - ⋆ sensor/actuator placement
  - ⋆ distributed estimation/control

non-smooth composite optimization

The two pillars

Proximal operator

$$\mathbf{prox}_{\mu g}(v) \;:=\; \underset{x}{\operatorname{argmin}} \;\; g(x) \;+\; \frac{1}{2\mu}\,\|x \,-\, v\|^2$$

Moreau envelope

$$M_{\mu g}(v) \;:=\; g(\mathbf{prox}_{\mu g}(v)) \;+\; \frac{1}{2\mu}\,\|\mathbf{prox}_{\mu g}(v) \,-\, v\|^2$$

The two pillars

Proximal operator

$$\mathbf{prox}_{\mu g}(v) \;:=\; \underset{x}{\operatorname{argmin}} \;\; g(x) \;+\; \frac{1}{2\mu} \|x \,-\, v\|^2$$

Moreau envelope

$$M_{\mu g}(v) \;:=\; g(\mathbf{prox}_{\mu g}(v)) \;+\; \frac{1}{2\mu} \|\mathbf{prox}_{\mu g}(v) \,-\, v\|^2$$

- **continuously differentiable**

  even when g is not

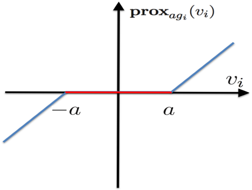
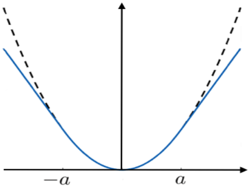
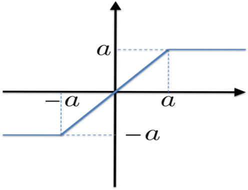
$$\nabla M_{\mu g}(v) \;=\; \frac{1}{\mu} \big(v \,-\, \mathbf{prox}_{\mu g}(v)\big)$$

Parikh & Boyd, FnT in Optimization ’14

Prox for ℓ_1 norm

$$\underset{x_i}{\text{minimize}} \quad \sum_i \left(\gamma \, |x_i| + \frac{1}{2\mu} (x_i - v_i)^2 \right)$$

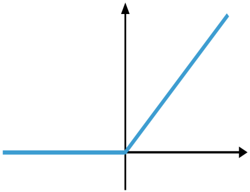
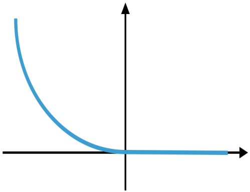
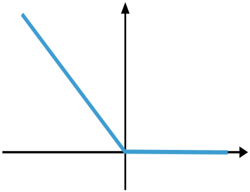
separability $\Rightarrow$ **element-wise analytical solution**

prox operator soft-thresholding	**Moreau envelope** Huber function	∇M saturation
		
	$a = \gamma\mu$	

Prox for $I_+(\cdot)$

$$\underset{x_i}{\text{minimize}} \quad \sum_i \left(I_+(x_i) \; + \; \frac{1}{2\mu} \left(x_i \; - \; v_i\right)^2 \right)$$

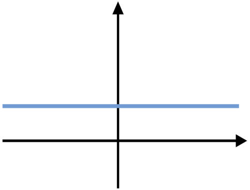
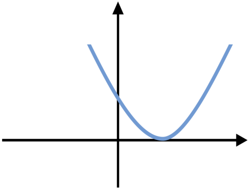
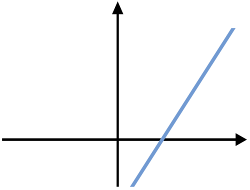
separability $\Rightarrow$ **element-wise analytical solution**

prox operator	**Moreau envelope**	∇M
		

Prox for $I_b(\cdot)$

$$\underset{x_i}{\text{minimize}} \quad \sum_i \left(I_{b_i}(x_i) + \frac{1}{2\mu} (x_i - v_i)^2 \right)$$

separability $\Rightarrow$ **element-wise analytical solution**

prox operator	**Moreau envelope**	∇M
		

Proximal gradient method

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

Generalizes gradient descent

$$x^{k+1} \; = \; \mathbf{prox}_{\alpha_k g}\big(x^k \; - \; \alpha_k \nabla f(x^k)\big)$$

Proximal gradient method

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

Generalizes gradient descent

$$x^{k+1} = \mathbf{prox}_{\alpha_k g}\big(x^k - \alpha_k \nabla f(x^k)\big)$$

- f convex; Lipschitz cts gradient $\Rightarrow$ convergence
- if $\mathbf{prox}_g$ easy to compute $\Rightarrow$ simple implementation
- cannot be applied to $g(\mathcal{T}(x))$
- acceleration with constraints (e.g., stability) challenging

Beck & Teboulle, SIAM J. Imaging Sci. '08

Augmented Lagrangian

Auxiliary variable

$$\begin{array}{ll} \underset{x,\,z}{\text{minimize}} & f(x) + g(z) \\ \text{subject to} & \mathcal{T}(x) - z = 0 \end{array}$$

- **benefit: decouples f and g**

Augmented Lagrangian

$$\mathcal{L}_\mu(x, z; y) = f(x) + g(z) + \langle y, \mathcal{T}(x) - z \rangle + \frac{1}{2\mu} \|\mathcal{T}(x) - z\|^2$$

Alternating Direction Method of Multipliers

$$x^{k+1} = \operatorname*{argmin}_{x} \ \mathcal{L}_\mu(x,\ z^k;\ y^k) \qquad \text{differentiable}$$

$$z^{k+1} = \operatorname*{argmin}_{z} \ \mathcal{L}_\mu(x^{k+1},\ z;\ y^k) \qquad \mathbf{prox}_{\mu g}(\cdot)$$

$$y^{k+1} = y^k + \frac{1}{\mu}\left(\mathcal{T}(x^{k+1}) - z^{k+1}\right)$$

- convenient for distributed implementation
- convergence speed: influenced by μ
- convergence for nonconvex f: active topic

Hong, Luo, Razaviyayn, SIOPT '16

Outline

- **Proximal augmented Lagrangian**
  - ⋆ continuously differentiable (even for non-smooth problems)

- **First-order primal-dual updates**

  Method of multipliers

  - ⋆ nonconvex f: convergence to a local minimum

  Arrow-Hurwicz-Uzawa gradient flow

  - ⋆ convenient for distributed optimization
  - ⋆ global exponential stability for strongly cvx problems

- **Second-order primal-dual updates**
  - ⋆ efficiently computable (e.g., for separable g)
  - ⋆ outstanding practical performance

Augmented Lagrangian

$$\begin{aligned} \underset{x,\,z}{\text{minimize}} \quad & f(x) \;+\; g(z) \\ \text{subject to} \quad & \mathcal{T}x \;-\; z \;=\; 0 \end{aligned}$$

Augmented Lagrangian

$$\mathcal{L}_\mu(x, z; y) \;=\; f(x) \;+\; g(z) \;+\; \langle y, \mathcal{T}x \;-\; z \rangle \;+\; \frac{1}{2\mu}\,\|\mathcal{T}x \;-\; z\|^2$$

Proximal augmented Lagrangian

Complete squares

$$\mathcal{L}_\mu(x, z; y) \;=\; f(x) \;+\; \underbrace{g(z) \;+\; \frac{1}{2\mu}\,\|z \,-\, (\mathcal{T}x \,+\, \mu y)\|^2}\; -\; \frac{\mu}{2}\,\|y\|^2$$

Proximal augmented Lagrangian

Complete squares

$$\mathcal{L}_\mu(x,z;y) \;=\; f(x) \;+\; \underbrace{g(z) \;+\; \frac{1}{2\mu}\left\|z \,-\, (\mathcal{T}x \,+\, \mu y)\right\|^2}\; - \;\frac{\mu}{2}\left\|y\right\|^2$$

Minimize over z

$$z_\mu^\star(x,y) \;=\; \mathbf{prox}_{\mu g}(\mathcal{T}x \,+\, \mu y)$$

Evaluate $\mathcal{L}_\mu$ at $z_\mu^\star$

$$\begin{aligned}\mathcal{L}_\mu(x;y) \;&:=\; \mathcal{L}_\mu(x,\, z_\mu^\star(x,y);\, y)\\ &=\; f(x) \,+\, M_{\mu g}(\mathcal{T}x \,+\, \mu y) \,-\, \frac{\mu}{2}\left\|y\right\|^2\end{aligned}$$

continuously differentiable

Forward-backward envelope

$$\begin{aligned}\mathcal{L}_{\mathrm{FBE}}(x) &:= \mathcal{L}_\mu(x; y = -\nabla f(x)) \\ &= f(x) + M_{\mu g}(x - \mu \nabla f(x)) - \frac{\mu}{2}\|\nabla f(x)\|^2\end{aligned}$$

Patrinos, Stella, Bemporad, arXiv:1402.6655

Forward-backward envelope

$$\begin{aligned} \mathcal{L}_{\mathrm{FBE}}(x) &:= \mathcal{L}_\mu(x; y = -\nabla f(x)) \\ &= f(x) + M_{\mu g}(x - \mu \nabla f(x)) - \frac{\mu}{2} \|\nabla f(x)\|^2 \end{aligned}$$

Patrinos, Stella, Bemporad, arXiv:1402.6655

Proximal gradient

- variable-metric gradient method on FBE

$$x^{k+1} = x^k - \mu \left(I - \mu \nabla^2 f(x)\right)^{-1} \nabla \mathcal{L}_{\mathrm{FBE}}(x)$$

OBJECTIVE

compute saddle points of the proximal augmented Lagrangian

Method of multipliers

$$
\begin{aligned}
(x^{k+1}, z^{k+1}) &= \underset{x,\,z}{\operatorname{argmin}}\ \mathcal{L}_{\mu_k}(x, z; y^k) \\
y^{k+1} &= y^k + \frac{1}{\mu_k}\left(\mathcal{T}(x^{k+1}) - z^{k+1}\right)
\end{aligned}
$$

Method of multipliers

$$x^{k+1} = \underset{x}{\operatorname{argmin}} \ \mathcal{L}_{\mu_k}(x; y^k)$$

$$y^{k+1} = y^k + \frac{1}{\mu_k}\left(\mathcal{T}(x^{k+1}) - z_\mu^\star(x^{k+1}, y^k)\right)$$

- nonconvex f: convergence to a local minimum
- x-minimization: differentiable problem

  e.g., can use L-BFGS
- adaptive μ-update

Arrow-Hurwicz-Uzawa gradient flow

Primal-descent Dual-ascent

$$\begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} = \begin{bmatrix} -\nabla_x \mathcal{L}_\mu(x;y) \\ \nabla_y \mathcal{L}_\mu(x;y) \end{bmatrix}$$

- continuous rhs even for non-differentiable g
- convenient for distributed implementation
- existing methods use subgradients

Feijer & Paganini, Automatica '10

Wang & Elia, CDC '11

Cherukuri, Gharesifard, Cortés, SICON '17

Primal-dual updates

$$\begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} = \begin{bmatrix} -\left(\nabla f(x) \; + \; T^T \nabla M_{\mu g}(Tx \; + \; \mu y)\right) \\ \mu \nabla M_{\mu g}(Tx \; + \; \mu y) \; - \; \mu y \end{bmatrix}$$

$$\mu \nabla M_{\mu g}(v) \; = \; v \; - \; \mathbf{prox}_{\mu g}(v)$$

- **Distributed implementation**
  - ⋆ g separable
  - ⋆ $\nabla f \colon \mathbb{R}^n \to \mathbb{R}^n$ sparse mapping
  - ⋆ $T^T T$ sparse matrix

Global exponential stability

$$\left.\begin{array}{rcl} g & - & \text{convex} \\ f & - & m_f \text{ strongly convex} \\ \nabla f & - & L_f \text{ Lipschitz cts} \\ TT^T & - & \text{full rank} \end{array}\right\} \Rightarrow \begin{array}{l} \text{there is } \rho > 0 \text{ s.t.} \\ \|\tilde{w}(t)\| \leq \alpha\, \mathrm{e}^{-\rho t} \|\tilde{w}(0)\| \end{array}$$

Dhingra, Khong, Jovanović, IEEE TAC '18
arXiv:1610.04514

Enabling tool

⋆ theory of **integral quadratic constraints**

Megretski & Rantzer, IEEE TAC '97

Lessard, Recht, Packard, SIOPT '16

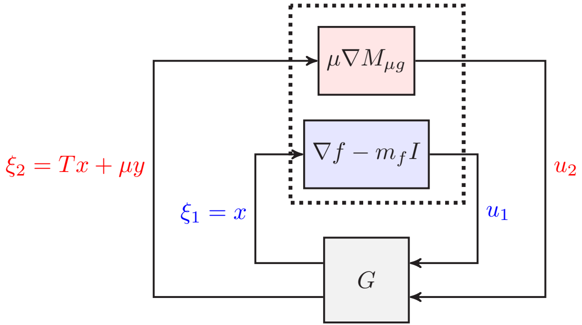
Hu, PhD Thesis '16

Hu, Seiler, Rantzer, COLT '17

System-theoretic viewpoint

$$\begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} = -\begin{bmatrix} m_f x \\ \mu y \end{bmatrix} - \begin{bmatrix} \nabla f(x) - m_f x \\ 0 \end{bmatrix} - \begin{bmatrix} T^T \nabla M_{\mu g}(Tx + \mu y) \\ -\mu \nabla M_{\mu g}(Tx + \mu y) \end{bmatrix}$$

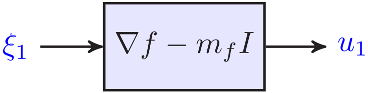
- stable linear system G in feedback with nonlinear terms

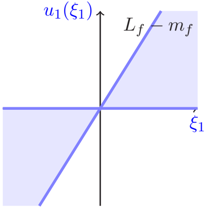


$$u_1(\xi_1) = \nabla f(\xi_1) - m_f \xi_1 \qquad u_2(\xi_2) = \xi_2 - \mathbf{prox}_{\mu g}(\xi_2)$$

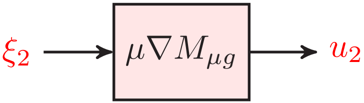
Nonlinearities

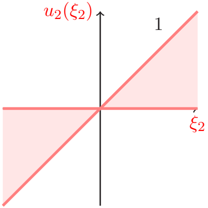
- **Gradient of convex function** $f(x) - \frac{m_f}{2}\|x\|^2$



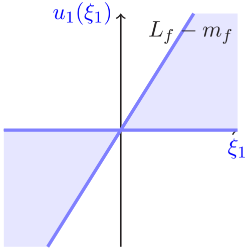


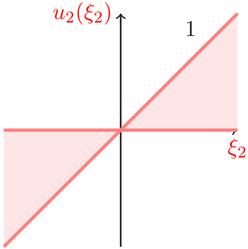
- **Scaled gradient of Moreau envelope**





Quadratic constraints





Pointwise quadratic inequalities

$$\begin{bmatrix} \xi_i - \xi_i^\star \\ u_i - u_i^\star \end{bmatrix}^T \underbrace{\begin{bmatrix} 0 & L_i\, I \\ L_i\, I & -2\, I \end{bmatrix}}_{\Pi_i} \begin{bmatrix} \xi_i - \xi_i^\star \\ u_i - u_i^\star \end{bmatrix} \geq 0$$

Lessard, Recht, Packard, SIOPT '16

- **KYP Lemma** Rantzer, SCL '96

$$\left.\begin{array}{c} \text{if there is } \rho > 0 \text{ s.t. } \forall\, \omega \in \mathbb{R} \\ \begin{bmatrix} G_\rho(\mathrm{j}\omega) \\ I \end{bmatrix}^* \Pi \begin{bmatrix} G_\rho(\mathrm{j}\omega) \\ I \end{bmatrix} \preceq 0 \end{array}\right\} \Rightarrow \|\tilde{w}(t)\| \leq \alpha\, \mathrm{e}^{-\rho t} \|\tilde{w}(0)\|$$

$$G_\rho(\mathrm{j}\omega) := C\,(\mathrm{j}\omega I - (A + \rho I))^{-1} B$$

Π – describes IQCs for u_1 and u_2

- **KYP Lemma** Rantzer, SCL '96

$$\left.\begin{array}{c} \text{if there is } \rho > 0 \text{ s.t. } \forall\, \omega \in \mathbb{R} \\ \begin{bmatrix} G_\rho(\mathrm{j}\omega) \\ I \end{bmatrix}^* \Pi \begin{bmatrix} G_\rho(\mathrm{j}\omega) \\ I \end{bmatrix} \preceq 0 \end{array}\right\} \Rightarrow \; \|\tilde{w}(t)\| \leq \alpha\, \mathrm{e}^{-\rho t} \|\tilde{w}(0)\|$$

$$G_\rho(\mathrm{j}\omega) := C\,(\mathrm{j}\omega I - (A + \rho I))^{-1} B$$

Π – describes IQCs for u_1 and u_2

⋆ **take Schur complement and diagonalize** TT^T

$$\omega^4 + b_i(\rho)\,\omega^2 + c_i(\rho) > 0$$

$$\begin{aligned} b_i &\;-\; \text{quadratic in } \rho \\ c_i &\;-\; \text{quartic in } \rho \\ b_i(0),\, c_i(0) &\;>\; 0 \end{aligned}$$

Example: distributed optimization

$$\left.\begin{array}{rl} \text{minimize} & \sum f_i(x_i) \\ \text{subject to} & Tx \;=\; 0 \end{array}\right\} \;\Leftrightarrow\; \text{minimize} \;\; \sum f_i(x_i) \;+\; g(Tx)$$

- ⋆ T^T – incidence matrix of a connected undirected network
- ⋆ $g(z) := \begin{cases} 0, & z = 0 \\ \infty, & z \neq 0 \end{cases}$

Example: distributed optimization

$$\left.\begin{array}{rl} \text{minimize} & \sum f_i(x_i) \\ \text{subject to} & Tx \;=\; 0 \end{array}\right\} \;\Leftrightarrow\; \text{minimize} \;\; \sum f_i(x_i) \; + \; g(Tx)$$

- ⋆ T^T – incidence matrix of a connected undirected network
- ⋆ $g(z) \;:=\; \begin{cases} 0, & z \;=\; 0 \\ \infty, & z \;\neq\; 0 \end{cases}$

Gradient flow dynamics

$$\begin{aligned} \dot{x} \;&=\; -\left(\nabla f(x) \;+\; (1/\mu)\, L\, x \;+\; \bar{y}\right) \\ \dot{\bar{y}} \;&=\; \beta L\, x \end{aligned}$$

- ⋆ each node stores $(x_i, \bar{y}_i)$ and communicates across $L \;:=\; T^T T$
- ⋆ $\bar{y} \;:=\; T^T y \;\Rightarrow\; \bar{y} \in \operatorname{span}\{\mathbb{1}^\perp\}$

- Forward-Euler discretization

$$\begin{aligned} x^{k+1} &= (I - (\alpha/\mu)L)\,x^k - \alpha\,\nabla f(x^k) - \alpha\,\bar{y}^k \\ \bar{y}^{k+1} &= \bar{y}^k + \alpha\beta L x^k \end{aligned}$$

EXTRA Algorithm

$$x^{k+1} = W x^k - \alpha\,\nabla f(x^k) + \frac{1}{2}\sum_{i=0}^{k-1}(W - I)\,x^i$$

Shi, Ling, Wu, Yin, SIOPT '15

follows from primal-dual gradient flow dynamics

326 citations on Google Scholar

Comments

- **Convex** f
  - ⋆ Lyapunov function

$$V = \frac{1}{2}\|x - x^{\star}\|^2 + \frac{1}{2}\|y - y^{\star}\|^2$$

  - ⋆ LaSalle's invariance principle: global asymptotic stability
  - ⋆ convergence rate?

- **Can handle multiple regularizers**

$$\text{minimize} \quad f(x) + \sum_i g_i(\mathcal{T}_i(x))$$

Rewrite as

$$\begin{aligned} \text{minimize} \quad & f(x) + \sum_i g_i(z_i) \\ \text{subject to} \quad & \mathcal{T}_i(x_i) - z_i = 0 \end{aligned}$$

⋆ convergence rate?

Second-order method of multipliers

Second-order updates

- f – strongly convex; twice cts differentiable
- $\mathbf{prox}_g$ – semismooth
- $T \in \mathbb{R}^{m\times n}$ – full row rank matrix

$$\nabla \mathcal{L}_\mu(x;y) = \begin{bmatrix} \nabla f(x) + \frac{1}{\mu} T^T(Tx + \mu y - \mathbf{prox}_{\mu g}(Tx + \mu y)) \\ Tx - \mathbf{prox}_{\mu g}(Tx + \mu y) \end{bmatrix}$$

P – B-subdifferential of $\mathbf{prox}_{\mu g}$

$$\partial_P^2 \mathcal{L}_\mu := \begin{bmatrix} \nabla^2 f + \frac{1}{\mu} T^T(I - P)T & T^T(I - P) \\ (I - P)T & -\mu P \end{bmatrix}$$

n negative and m positive e-values

Second-order updates

- f – strongly convex; twice cts differentiable
- $\mathbf{prox}_g$ – semismooth
- $T \in \mathbb{R}^{m \times n}$ – full row rank matrix

$$\nabla \mathcal{L}_\mu(x; y) = \begin{bmatrix} \nabla f(x) + \frac{1}{\mu} T^T (Tx + \mu y - \mathbf{prox}_{\mu g}(Tx + \mu y)) \\ Tx - \mathbf{prox}_{\mu g}(Tx + \mu y) \end{bmatrix}$$

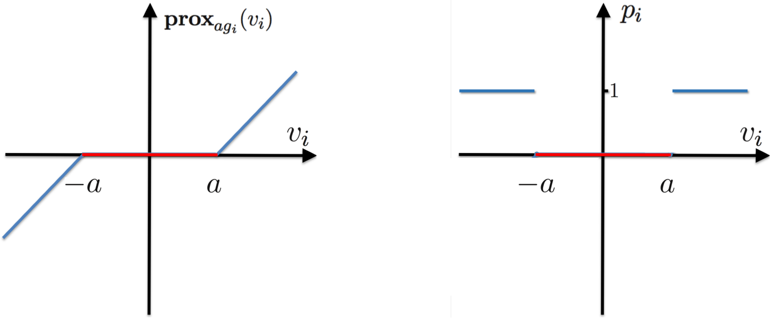
P – B-subdifferential of $\mathbf{prox}_{\mu g}$

$$\partial_P^2 \mathcal{L}_\mu := \begin{bmatrix} \nabla^2 f + \frac{1}{\mu} T^T (I - P) T & T^T (I - P) \\ (I - P) T & -\mu P \end{bmatrix}$$

n negative and m positive e-values

separable g $\Rightarrow$ diagonal P

- **Example: ℓ_1 norm**



soft-thresholding

$$p_i \in \begin{cases} 0 & |v_i| < a \\ 1 & |v_i| > a \\ \{0,1\} & |v_i| = a \end{cases}$$

Continuous-time dynamics

Differential inclusion

$$\begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} \in -(\partial_C^2 \mathcal{L}_\mu(x;y))^{-1} \nabla \mathcal{L}_\mu(x;y)$$

- saddle points of $\mathcal{L}_\mu(x;y)$
  - ⋆ globally exponentially stable
  - ⋆ Lyapunov function $\|\nabla \mathcal{L}_\mu(x;y)\|^2$

Second-order update

$$\begin{bmatrix} x^{k+1} \\ y^{k+1} \end{bmatrix} = \begin{bmatrix} x^{k} \\ y^{k} \end{bmatrix} - \alpha_k \begin{bmatrix} \tilde{x}_k \\ \tilde{y}_k \end{bmatrix}$$

Search direction

$$\partial_P^2 \mathcal{L}_\mu(x^k; y^k) \begin{bmatrix} \tilde{x}_k \\ \tilde{y}_k \end{bmatrix} = -\nabla \mathcal{L}_\mu(x^k; y^k)$$

Key challenge

- **How to assess progress?**

Merit function

$$\mathcal{M}_\mu(x, z; y, y_\mathrm{e}) \;:=\; \mathcal{L}_\mu(x, z; y_\mathrm{e}) \;+\; \frac{1}{2\mu}\,\|Tx \,-\, z \,+\, \mu(y_\mathrm{e} \,-\, y)\|^2$$

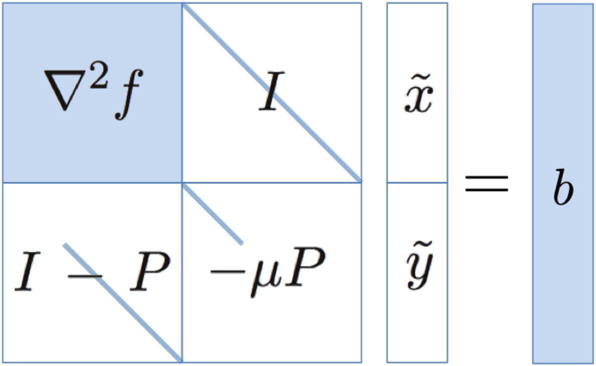
y_e – Lagrange multiplier estimate

step-size selection: backtracking

Gill & Robinson, Comput. Optim. Appl. '12

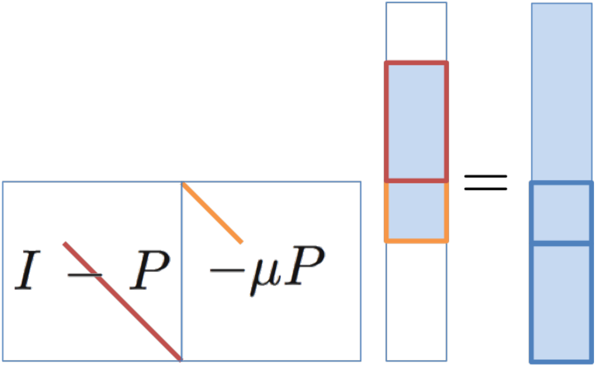
Efficient 2nd-order update?

Example: $T = I;\ g(x) = \|x\|_1 \ \Rightarrow\ p_i \in \{0, 1\}$



Efficient 2nd-order update?

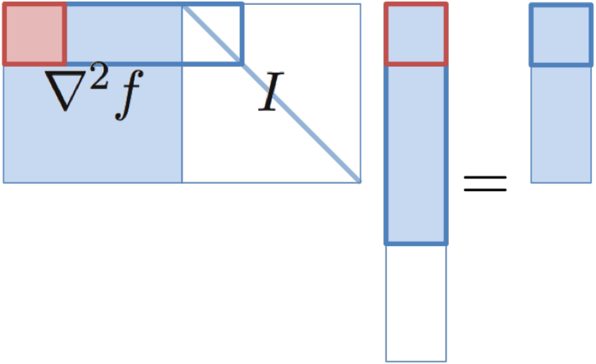
Example: $T = I;\ g(x) = \|x\|_1 \ \Rightarrow\ p_i \in \{0, 1\}$



- explicit evaluation

Efficient 2nd-order update?

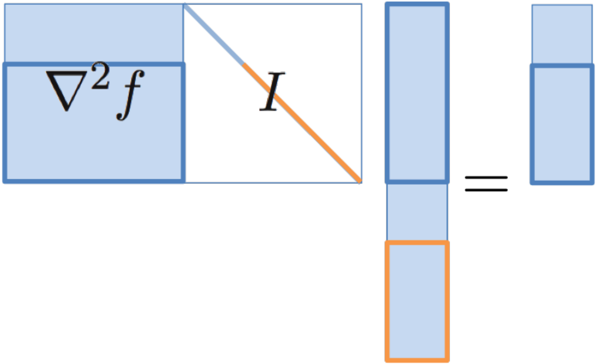
Example: $T = I;\ g(x) = \|x\|_1 \ \Rightarrow\ p_i \in \{0, 1\}$



- **limited matrix inversion** (independent of μ)

Efficient 2nd-order update?

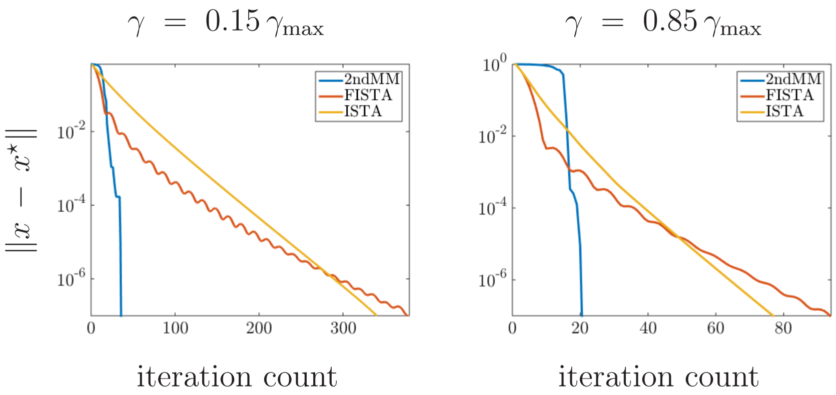
Example: $T = I;\ g(x) = \|x\|_1 \Rightarrow p_i \in \{0, 1\}$



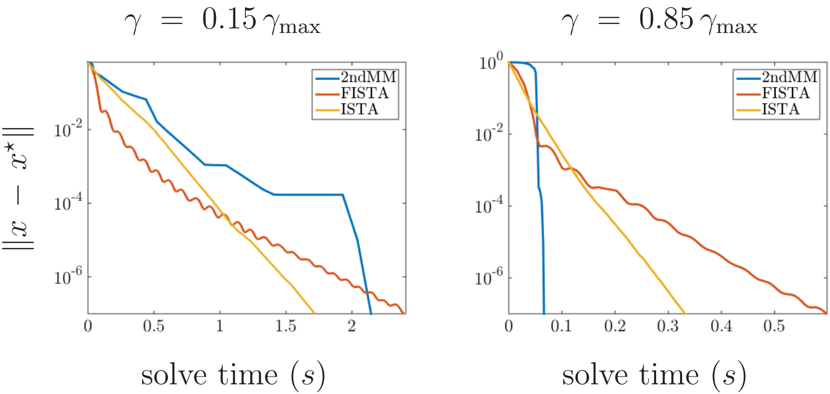
- **matrix-vector multiplication**

Computational experiments: LASSO

$$\text{minimize} \quad \frac{1}{2}\|Ax - b\|^2 + \gamma\|x\|_1$$

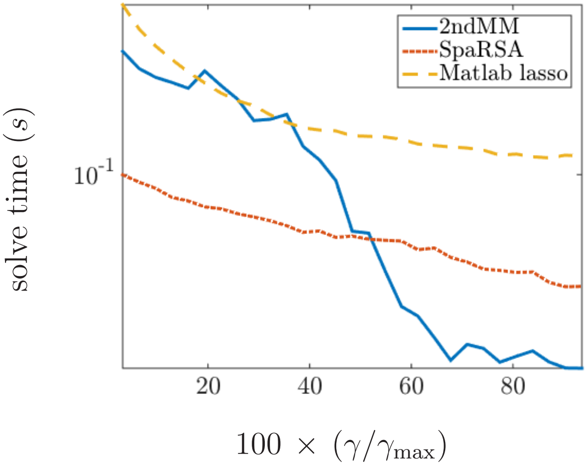


$$\text{minimize} \quad \frac{1}{2}\|Ax - b\|^2 + \gamma\|x\|_1$$



Influence of γ

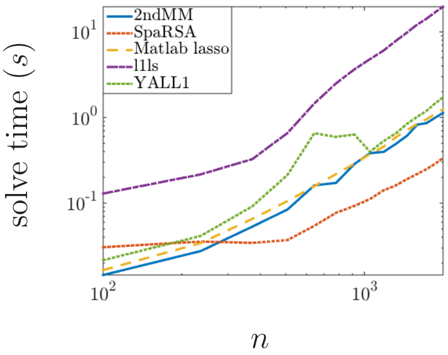
$$\text{minimize} \quad \frac{1}{2}\|Ax - b\|^2 + \gamma\|x\|_1$$

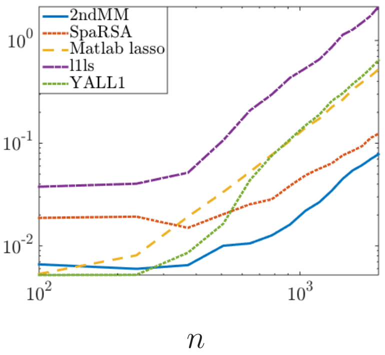


Influence of problem size

$$\text{minimize} \quad \frac{1}{2}\|Ax - b\|^2 + \gamma\|x\|_1$$

$\gamma = 0.15\,\gamma_{\max}$ $\gamma = 0.85\,\gamma_{\max}$





$x \in \mathbb{R}^n;\ A \in \mathbb{R}^{2n\times n}$

Remarks

- **Proximal augmented Lagrangian**
  - ⋆ Continuously differentiable (even for non-smooth problems)
  - ⋆ Method of multipliers
  - ⋆ Arrow-Hurwicz-Uzawa dynamics
- **Second-order updates**
  - ⋆ Efficiently computable (e.g., for separable g and large γ)
  - ⋆ Good practical performance
- **Challenges**
  - ⋆ Establishing convergence rate without strong convexity
  - ⋆ Alternative merit functions
  - ⋆ Convergence for nonconvex problems

References

First-order methods

- ⋆ Dhingra, Khong, Jovanović, IEEE TAC ’18
- ⋆ Ding, Hu, Dhingra, Jovanović, CDC ’18

Second-order methods

- ⋆ Dhingra, Khong, Jovanović, arXiv:1709.01610

Distributed method for non-smooth problems

- ⋆ Hassan-Moghaddam & Jovanović, ACC ’18

Lyapunov-based characterization

- ⋆ Qu & Li, IEEE CSL ’19; Ding & Jovanović, ACC ’19
- ⋆ Cherukuri, Mallada, Low, Cortes, IEEE TAC ’18

Feedback-based online optimization

- ⋆ Colombino, Dall’Anese, Bernstein, arXiv:1805.09877