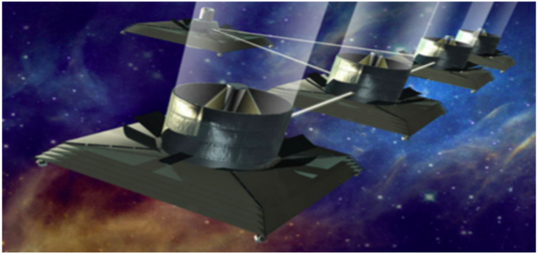
Proximal method of multipliers algorithm for sparsity-promoting optimal control

Neil K. Dhingra and Mihailo R. Jovanović

UNIVERSITY OF MINNESOTA

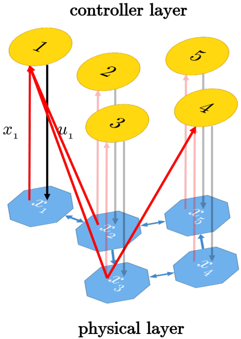
2016 American Control Conference, Boston, MA

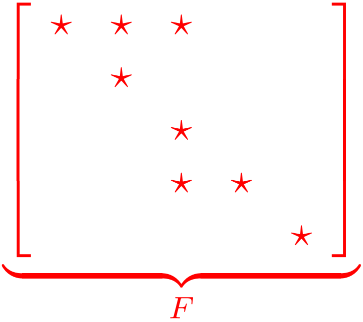
Large-scale distributed systems



Sparsity-promoting state-feedback design

$$\begin{aligned} \dot{x} &= (A - B_2 F)\, x + B_1 d \\ z &= \begin{bmatrix} Q^{1/2} \\ -R^{1/2} F \end{bmatrix} x \end{aligned}$$





Sparsity pattern of F determines communication topology

Sparsity-promoting optimal control

$$\text{minimize} \quad J(F) \quad + \quad \gamma \sum_{ij} |F_{ij}|$$

$$\downarrow \qquad\qquad\qquad \downarrow$$

closed-loop $\mathcal{H}_2$ norm — **sparsity-promoting penalty function**

(Lin, Fardad, Jovanović TAC '13), (Fardad, Lin, Jovanović ACC '11)

Sparsity-promoting optimal control

$$\text{minimize} \quad J(F) \quad + \quad \gamma \sum_{ij} |F_{ij}|$$

$J(F)$ ↓ **closed-loop $\mathcal{H}_2$ norm**

$\gamma \sum_{ij} |F_{ij}|$ ↓ **sparsity-promoting penalty function**

$\gamma > 0$ specifies tradeoff

- $\gamma = 0$ — Linear Quadratic Regulator: dense, centralized
- $\gamma \uparrow$ — sparser F: distributed

(Lin, Fardad, Jovanović TAC ‘13), (Fardad, Lin, Jovanović ACC ‘11)

Related work

- Sparsity-promoting $\mathcal{H}_\infty$ control
  - (Schuler, Li, Lam, Allgöwer, IJC '11)
  - (Schuler, Münz, Allgöwer, IFAC '12)
- Convex relaxations
  - (Lavaei, Allerton '13)
  - (Fazelnia, Madani, Lavaei, CDC '14)
  - (Fardad, Jovanović, ACC '14)
- Atomic norm regularization
  - (Matni, CDC '13; TCNS '16)
  - (Matni and Chandrasekaran, TAC '16)
  - (Wang, Matni, Doyle, CDC '14)

Structured synthesis via regularization

$$\text{minimize} \quad \underbrace{J(F)}_{\text{performance}} \; + \; \gamma \, \underbrace{g(F)}_{\text{structure}}$$

minimize $J(F)$ + $\gamma\, g(F)$

↓ **performance**   ↓ **structure**

General challenges

- $J(F)$ nonconvex but differentiable
- $g(F)$ nondifferentiable but separable

Proximal Method of Multipliers

New framework for nonconvex regularized problems

- Guaranteed convergence to local minima
- General regularization $g(\mathcal{T}(F))$
- Second-order information, limited parameter tuning

Proximal Method of Multipliers

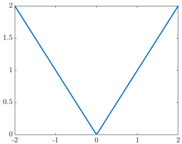
New framework for nonconvex regularized problems

- Guaranteed convergence to local minima
- General regularization $g(\mathcal{T}(F))$
- Second-order information, limited parameter tuning

This work

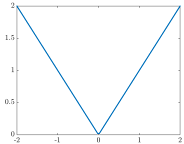
- Application to sparsity-promoting optimal control
- Comparison with proximal gradient, ADMM

Proximal operators

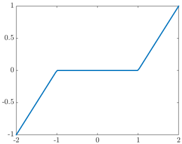


$$g = |\cdot|$$

Proximal operators



$g = |\cdot|$

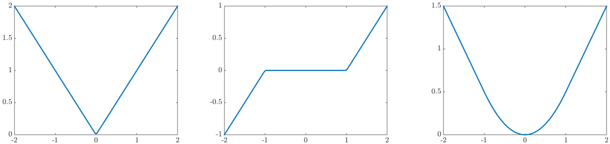


$\mathbf{prox}_g$(Soft-thresh)

Proximal operator

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

Proximal operators



$g = |\cdot|$ $\quad$ $\mathbf{prox}_g$(Soft-thresh) $\quad$ M_g (Huber)

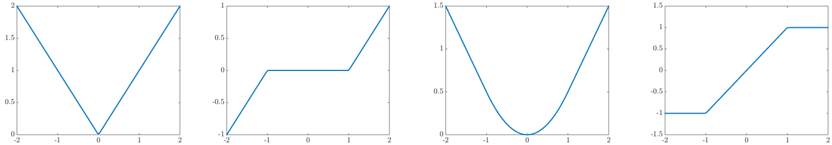
Proximal operator

$$\mathbf{prox}_{\mu g}(V) \;:=\; \operatorname*{argmin}_F \; g(F) + \frac{1}{2\mu}\|F - V\|_F^2$$

Moreau envelope

$$M_{\mu g}(V) \;:=\; \inf_F \; g(F) + \frac{1}{2\mu}\|F - V\|_F^2$$

Proximal operators



$g = |\cdot|$ $\quad$ $\mathbf{prox}_g$(Soft-thresh) $\quad$ M_g (Huber) $\quad$ ∇M_g (Saturation)

Proximal operator

$$\mathbf{prox}_{\mu g}(V) \;:=\; \operatorname*{argmin}_F \; g(F) \;+\; \frac{1}{2\mu}\,\|F \;-\; V\|_F^2$$

Moreau envelope

$$M_{\mu g}(V) \;:=\; \inf_F \; g(F) \;+\; \frac{1}{2\mu}\,\|F \;-\; V\|_F^2$$

Continuously differentiable gradient

(Parikh and Boyd, FnT Opt '14)

Proximal gradient descent

$$\text{minimize} \quad J(F) \; + \; \gamma \, g(F)$$

Generalization of gradient descent

$$F^{k+1} \; = \; \mathbf{prox}_{\gamma\alpha g}\left(F^{k} \; - \; \alpha\nabla J(F^{k})\right)$$

step-size α

Proximal gradient descent

$$\text{minimize} \quad J(F) \; + \; \gamma \, g(F)$$

Generalization of gradient descent

$$F^{k+1} \; = \; \mathbf{prox}_{\gamma\alpha g}\left(F^k \; - \; \alpha \nabla J(F^k)\right)$$

step-size α

- Simple if $\mathbf{prox}_g$ simple
- Cannot be applied to $g(\mathcal{T}(F))$
- Acceleration with constraints challenging (e.g., stability)

(Beck and Teboulle, SIAM JIS '08)

Alternative approach

Auxiliary variable **decouples** J and g

$$\begin{array}{ll} \underset{F,\,G}{\text{minimize}} & J(F) \;+\; \gamma\, g(G) \\ \text{subject to} & F \;-\; G \;=\; 0 \end{array}$$

Form augmented Lagrangian

$$\mathcal{L}_\mu(F,\, G; \Lambda) \;=\; J(F) \;+\; g(G) \;+\; \langle \Lambda, F \;-\; G \rangle \;+\; \frac{1}{2\mu} \| F \;-\; G \|_F^2$$

Alternating Direction Method of Multipliers (ADMM)

$$F^{k+1} = \underset{F}{\text{argmin}} \;\; \mathcal{L}_\mu(F,\, G^k;\, \Lambda^k) \qquad \text{differentiable}$$

$$G^{k+1} = \underset{G}{\text{argmin}} \;\; \mathcal{L}_\mu(F^{k+1},\, G;\, \Lambda^k) \qquad \mathbf{prox}_{\mu g}(\cdot)$$

$$\Lambda^{k+1} = \Lambda^k + \frac{1}{\mu}\,(F^{k+1} - G^{k+1})$$

(Lin, Fardad, Jovanović, TAC '13)

Alternating Direction Method of Multipliers (ADMM)

$$
\begin{aligned}
F^{k+1} &= \underset{F}{\operatorname{argmin}} \;\; \mathcal{L}_\mu(F,\, G^k;\, \Lambda^k) && \text{differentiable} \\
G^{k+1} &= \underset{G}{\operatorname{argmin}} \;\; \mathcal{L}_\mu(F^{k+1},\, G;\, \Lambda^k) && \mathbf{prox}_{\mu g}(\cdot) \\
\Lambda^{k+1} &= \Lambda^k \;+\; \frac{1}{\mu}\,(F^{k+1} \;-\; G^{k+1})
\end{aligned}
$$

- convergence not guaranteed for nonconvex problems
- speed sensitive to parameter μ

(Lin, Fardad, Jovanović, TAC '13)

Method of Multipliers

$$
\begin{aligned}
(F^{k+1}, G^{k+1}) &= \underset{F,\, G}{\operatorname{argmin}}\ \mathcal{L}_{\mu^k}(F, G; \Lambda^k) \\
\Lambda^{k+1} &= \Lambda^k + \frac{1}{\mu^k}(F^{k+1} - G^{k+1})
\end{aligned}
$$

- guaranteed convergence to a local minimum
- systematic procedure to adjust μ
- Joint (F, G) subproblem

Elimination of G

Rewrite $\mathcal{L}_\mu$

$$\mathcal{L}_\mu(F, G; \Lambda) = J(F) + \underbrace{\gamma g(G) + \|G - (F + \mu\Lambda)\|_F^2} - \frac{\mu}{2}\|\Lambda\|_F^2$$

Elimination of G

Rewrite $\mathcal{L}_\mu$

$$\mathcal{L}_\mu(F, G; \Lambda) = J(F) + \underbrace{\gamma g(G) + \|G - (F + \mu\Lambda)\|_F^2} - \frac{\mu}{2}\|\Lambda\|_F^2$$

$$G^* = \mathbf{prox}_{\gamma\mu g}(F + \mu\Lambda)$$

Elimination of G

Rewrite $\mathcal{L}_\mu$

$$\mathcal{L}_\mu(F,G;\Lambda) = J(F) + \underbrace{\gamma g(G) + \|G - (F + \mu\Lambda)\|_F^2}_{} - \frac{\mu}{2}\|\Lambda\|_F^2$$

$$G^* = \mathbf{prox}_{\gamma\mu g}(F + \mu\Lambda)$$

Eliminate G

$$\mathcal{L}_\mu(F;\Lambda) = J(F) + \gamma M_{\gamma\mu g}(F + \mu\Lambda) - \frac{\mu}{2}\|\Lambda\|_F^2$$

Proximal Method of Multipliers

$$
\begin{aligned}
F^{k+1} &= \underset{F}{\operatorname{argmin}}\ \mathcal{L}_{\mu^k}(F; \Lambda^k) \\
\Lambda^{k+1} &= \Lambda^k + \frac{1}{\mu^k}\left(F^{k+1} - \underbrace{\mathbf{prox}_{\mu\gamma g}(F^{k+1} + \mu\Lambda^k)}_{G^*}\right)
\end{aligned}
$$

Proximal Method of Multipliers subproblem

$$\underset{F}{\text{minimize}} \quad J(F) \; + \; \gamma \, M_{\gamma\mu^k g}(F \; + \; \mu^k \Lambda^k)$$

$M_{\mu g}$ separable if g separable

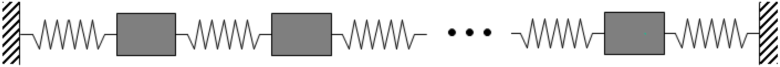
- Proximal gradient
- BB step-size selection

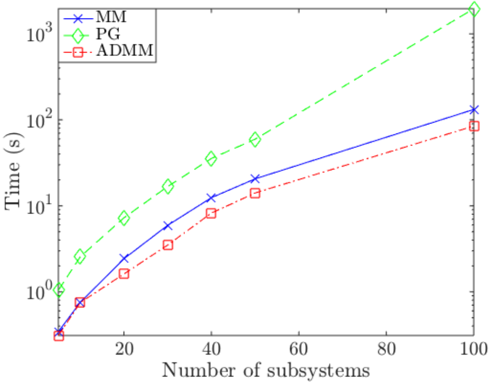
$\mathcal{L}_\mu$ **once** continuously differentiable

- quasi-Newton methods
- L-BFGS works well in practice (Lewis and Overton, '13)

Results: mass-spring

$$\ddot{x}_i = (x_{i+1} - x_i) + (x_{i-1} - x_i) + d_i + u_i$$

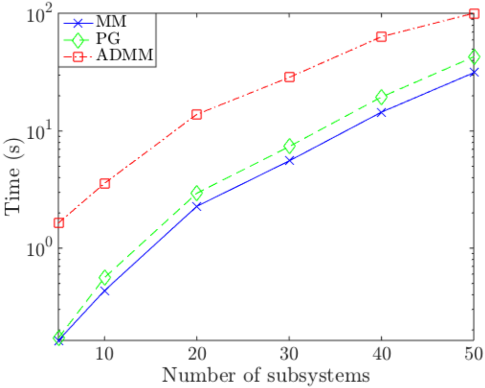




Results: unstable network

Unstable coupled nodes

$$\begin{bmatrix} \dot{x}_{1i} \\ \dot{x}_{2i} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} x_{1i} \\ x_{2i} \end{bmatrix} + \sum_{j \neq i} \mathrm{e}^{-\delta(i,j)} \begin{bmatrix} x_{1j} \\ x_{2j} \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} (d_i + u_i)$$



(Motee and Jadbabaie, TAC '08)

New framework for composite minimization

Relative to proximal gradient

- more general problems
- second-order information

Relative to ADMM

- convergence guarantee
- less parameter tuning
- comparable or better speed in practice

Acknowledgements

Support:

- NSF ECCS-1407958
- University of Minnesota DDF
- UMN Informatics Institute Transdisciplinary Faculty Fellowship