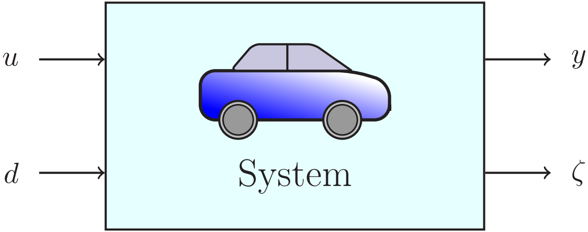
Optimization and control of large-scale networked systems

Neil K. Dhingra

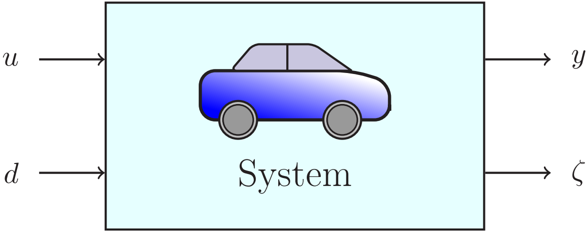
UNIVERSITY OF MINNESOTA

Final Oral Presentation
December 20, 2016

Feedback control – cruise control example

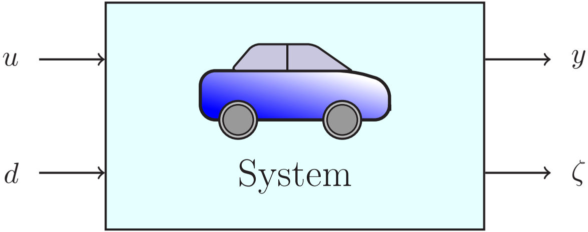


Feedback control – cruise control example



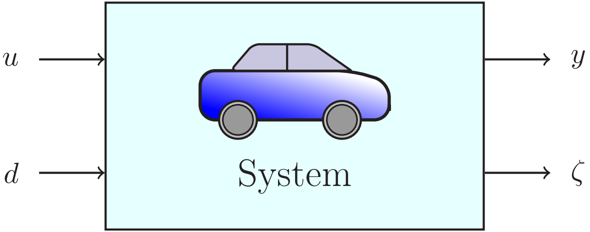
- u – control input (accelerator/brake)

Feedback control – cruise control example



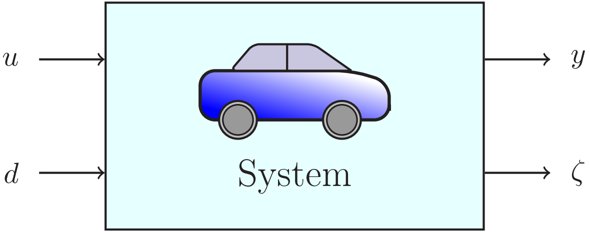
- u – control input (accelerator/brake)
- y – measured output (speedometer)

Feedback control – cruise control example



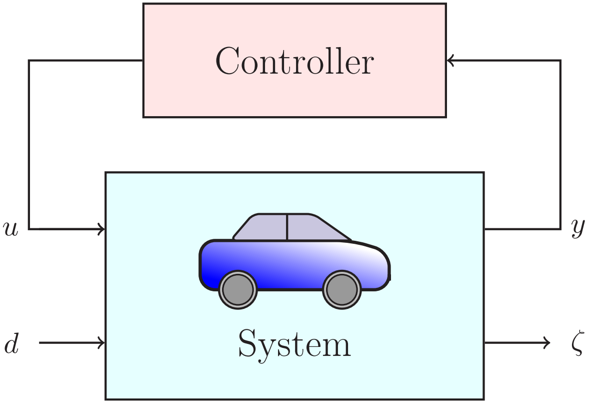
- u – control input (accelerator/brake)
- y – measured output (speedometer)
- d – disturbance input (hills, rain, etc.)

Feedback control – cruise control example



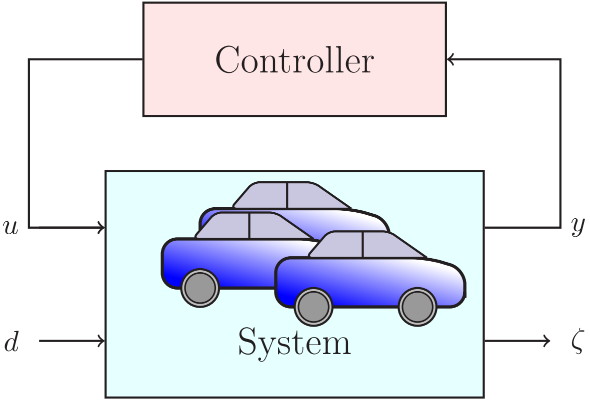
- u – control input (accelerator/brake)
- y – measured output (speedometer)
- d – disturbance input (hills, rain, etc.)
- ζ – regulated output (difference from desired speed)

Feedback control – cruise control example



- u – control input (accelerator/brake)
- y – measured output (speedometer)
- d – disturbance input (hills, rain, etc.)
- ζ – regulated output (difference from desired speed)

Networks of systems

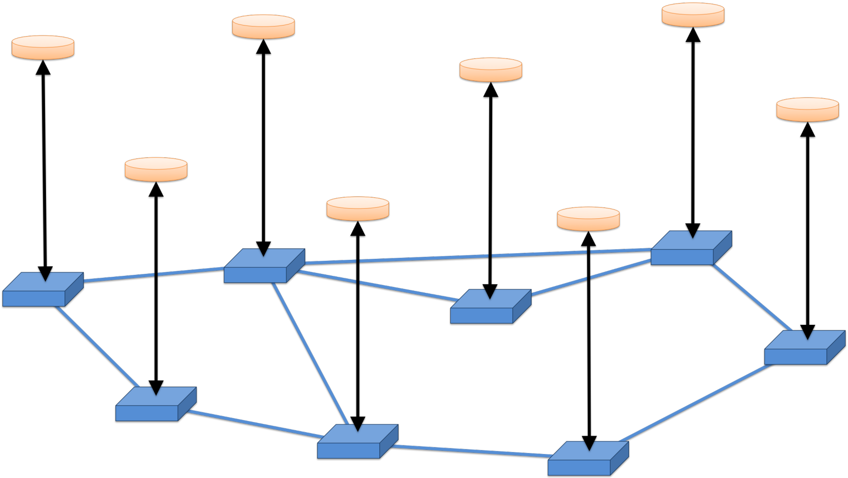


- Many distinct subsystems
- Traditional control – one *centralized* controller

Distributed control

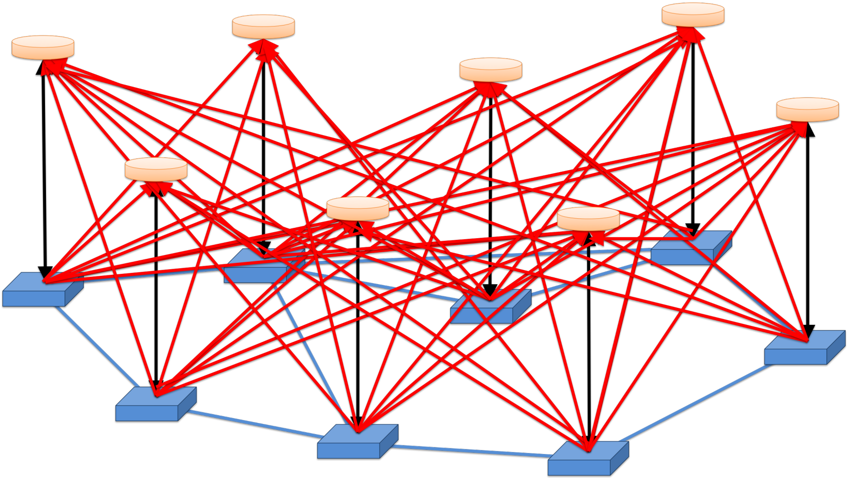
- Independent controllers must coordinate
- Here, simple communication topology

Structured control



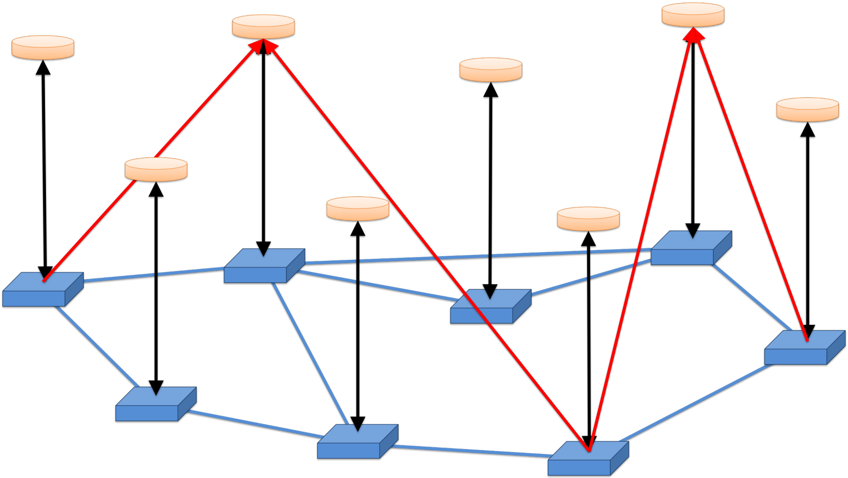
- Separate subsystems have associated controllers

Structured control



- Separate subsystems have associated controllers
- Traditional controller design: all-to-all communication architecture

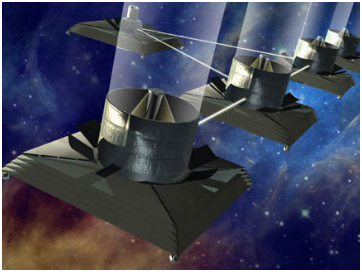
Structured control



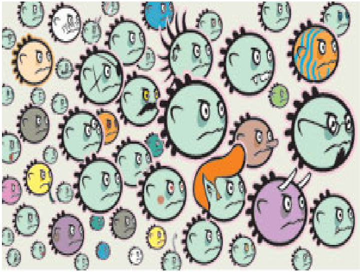
- Separate subsystems have associated controllers
- Want: sparse communication architecture

Applications

SATELLITE FORMATIONS

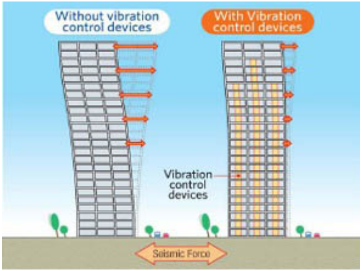


COMBINATION DRUG THERAPY



POWER NETWORKS

CONTROL OF BUILDINGS



Outline

I STRUCTURED OPTIMAL CONTROL

- regularization
- convex classes
- example: combination drug therapy

II NONCONVEX REGULARIZED PROBLEMS

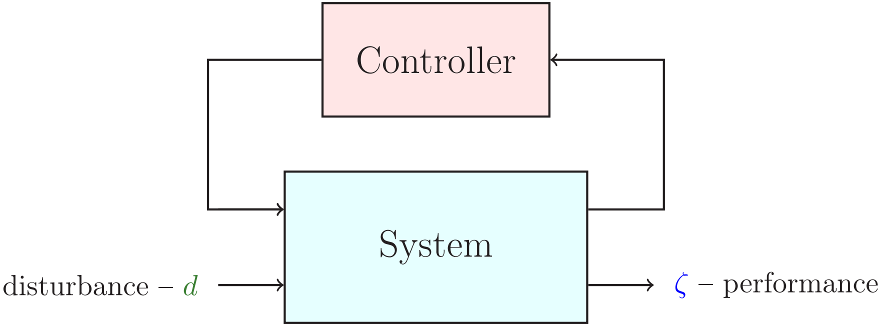
- *proximal* augmented Lagrangian
- example: edge addition in directed consensus networks

III CONVEX REGULARIZED PROBLEMS

- Second-order saddle point dynamics
- example: LASSO

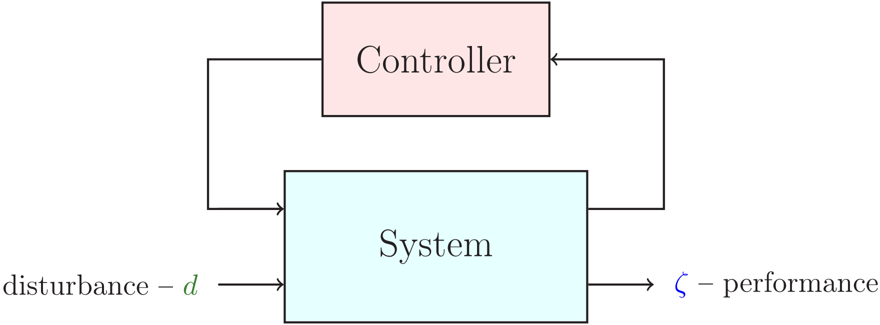
Structured optimal control

Effect of control



dynamics: $\dot{\psi} \;=\; (A \;-\; F(x))\,\psi \;+\; B_1\,d$

Effect of control



$$\text{dynamics:} \quad \dot{\psi} \;=\; (A \;-\; F(x))\,\psi \;+\; B_1\,d$$

$$\text{performance:} \quad \zeta \;=\; \begin{bmatrix} Q^{1/2} \\ 0 \end{bmatrix} \psi \;+\; \begin{bmatrix} 0 \\ D(x) \end{bmatrix} \psi$$

Static feedback

$$\begin{aligned}\dot{\psi} &= A\psi + B_1 d + B_2 u \\ \zeta &= \begin{bmatrix} Q^{1/2} \\ 0 \end{bmatrix}\psi + \begin{bmatrix} 0 \\ R^{1/2} \end{bmatrix} u\end{aligned}$$

STATE-FEEDBACK: $u = -X\psi$

- $F(X) = -B_2 X$
- $D(X) = -R^{1/2} X$

Static feedback

$$\begin{aligned}\dot{\psi} &= A\psi + B_1 d + B_2 u \\ \zeta &= \begin{bmatrix} Q^{1/2} \\ 0 \end{bmatrix}\psi + \begin{bmatrix} 0 \\ R^{1/2} \end{bmatrix} u\end{aligned}$$

STATE-FEEDBACK: $u = -X\psi$

- $F(X) = -B_2 X$
- $D(X) = -R^{1/2} X$

OUTPUT FEEDBACK: $y = C_2\psi$, $u = -Xy$

- $F(X) = -B_2 X C_2$
- $D(X) = -R^{1/2} X C_2$

Performance metrics

DESIGN **controller** x TO MINIMIZE **effect of d on ζ**

- CLOSED-LOOP $\mathcal{H}_2$ NORM

$$f_2(x) \; := \; \lim_{t \to \infty} \mathbb{E} \left(\zeta^T(t) \, \zeta(t) \right)$$

  - Variance amplification $d \to \zeta$
  - 'Average' measure

Performance metrics

DESIGN **controller** x TO MINIMIZE **effect of d on ζ**

- CLOSED-LOOP $\mathcal{H}_2$ NORM

$$f_2(x) \;:=\; \lim_{t \to \infty} \mathbb{E}\left(\zeta^T(t)\,\zeta(t)\right)$$

  - Variance amplification $d \to \zeta$
  - 'Average' measure

- CLOSED-LOOP $\mathcal{H}_\infty$ NORM

$$f_\infty(x) \;:=\; \sup_{\|d\|_{\mathcal{L}_2}=1} \frac{\|\zeta\|_{\mathcal{L}_2}}{\|d\|_{\mathcal{L}_2}}$$

  - Worst-case amplification $d \to \zeta$
  - Peak of frequency response

Structure via regularization

$$\text{minimize} \quad \underbrace{f(x)}_{\textbf{performance}} \quad + \quad \underbrace{\gamma\, g(Tx)}_{\textbf{structure}}$$

- f – possibly nonconvex; cts-differentiable
- g – convex; often non-differentiable
- $\gamma \geq 0$ – performance vs structure tradeoff

Structure via regularization

$$\text{minimize} \quad \underbrace{f(x)}_{\textbf{performance}} \quad + \quad \underbrace{\gamma\, g(Tx)}_{\textbf{structure}}$$

- f – possibly nonconvex; cts-differentiable
- g – convex; often non-differentiable
- $\gamma \geq 0$ – performance vs structure tradeoff

- x – design variable (e.g., feedback gain matrix)
- Tx – enforce structure in *image* of x

Typical regularizers

Examples

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

Typical regularizers

EXAMPLES

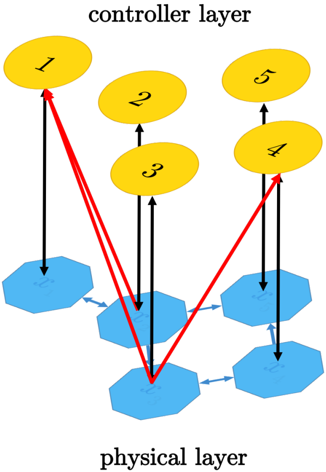
- $g(x) = \|x\|_1 = \sum |x_i|$ sparse x
- $g(x) = I_{\mathcal{C}}(x)$ convex constraints
- $g(X) = \|X\|_*$ low rank x

APPLICATIONS

- Compressed sensing LASSO
- Machine learning regularized logistic regression
- 'Netflix problem' matrix completion

Negahban, Yu, Wainwright, Ravikumar, '09

Limited communication via sparse feedback control



$$X = \begin{bmatrix} * & * & * & & \\ & * & & & \\ & & * & & \\ & & * & * & \\ & & & & * \end{bmatrix}$$

$$A = \begin{bmatrix} * & * & & & \\ * & * & & & * \\ & * & * & * & \\ & & * & * & \\ & * & & * & * \end{bmatrix}$$

Structured $\mathcal{H}_\infty$ control

KYP LEMMA: $f_\infty(x) \leq \lambda$ iff $\exists\, P \succ 0$ s.t.

$$\begin{bmatrix} C^TC + (A - F(x))^T P + P(A - F(x)) & PB \\ B^TP & -\lambda^2 I \end{bmatrix} \prec 0$$

Structured $\mathcal{H}_\infty$ control

KYP Lemma: $f_\infty(x) \leq \lambda$ iff $\exists\, P \succ 0$ s.t.

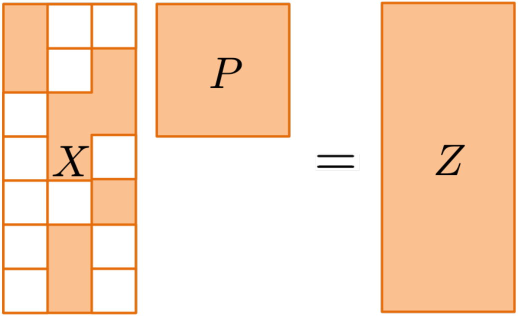
$$\begin{bmatrix} C^T C + (A - F(x))^T P + P(A - F(x)) & PB \\ B^T P & -\lambda^2 I \end{bmatrix} \prec 0$$

Convexity via $Z := PF(x)$

$$\begin{bmatrix} C^T C + A^T P - Z^T + PA - Z & PB \\ B^T P & -\lambda^2 I \end{bmatrix} \prec 0$$

Similar change of variables for $\mathcal{H}_2$

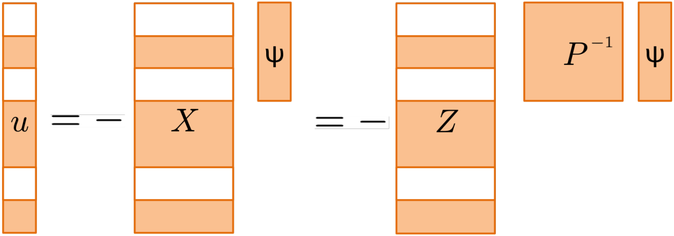
Structure via regularization



- State-feedback $F(X)P = -B_2XP = -B_2Z$
- Structural constraints on X bilinear in P, Z

Actuator/sensor selection

- Row-sparsity preserved
- Select inputs which yield best closed-loop performance

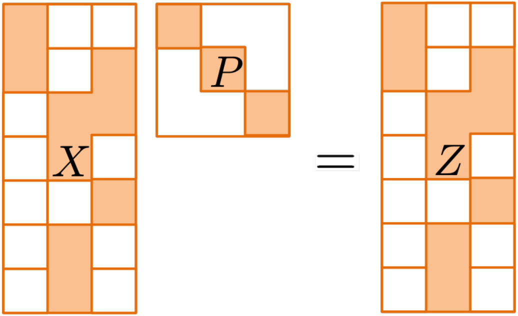


- Efficient algorithm for actuator/sensor selection SDPs

Dhingra, Jovanović, Luo, CDC '14

Structure via regularization

DIAGONAL P: preserve sparsity structure



- Conservative (except for $\mathcal{H}_\infty$ control of positive systems)
- Bilinear box constraints $X_{ij} = Z_{ij}/P_{jj} \in [-1, 1]$

Tanaka & Langbort, TAC '11
Rantzer '12, '15

Decentralized control of positive systems

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

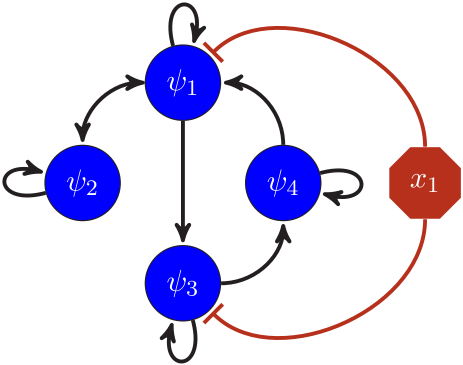
- Positive systems: $\psi(0) \geq 0$, $d(t) \geq 0$ implies $\psi(t) \geq 0$ for all t
  - $B_1, Q^{1/2} \geq 0$
  - A Metzler (off-diagonal elements nonnegative)
  - Decentralized control – $F(x)$ diagonal

Examples

- Leader selection in directed consensus networks
- Combination drug therapy

Combination drug therapy

$$\dot{\psi} = \left(A + \sum_{k=1}^{m} x_k F_k \right) \psi + d$$



$$\underbrace{\begin{bmatrix} \star & \star & & \star \\ \star & \star & & \\ \star & & \star & \\ & & \star & \star \end{bmatrix}}_{A}$$

$$\underbrace{\begin{bmatrix} \star & & & \\ & & & \\ & & \star & \\ & & & \end{bmatrix}}_{F_1}$$

- Mutant ψ_i mutates to ψ_j at rate A_{ji}
- Drug x_k kills ψ_i at rate $(F_k)_{ii}$

Rantzer & Bernhardsson CDC '14

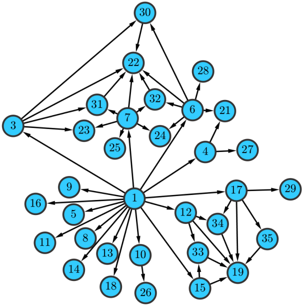
Jonsson, Rantzer, Matni, Murray CDC '14

Decentralized control of positive systems

- Background
  - Combination drug therapy — Jonsson et al CDC '14
  - $\mathcal{L}_1$ control — Rantzer & Bernhardsson CDC '14; Colaneri et al., AUT '14
- Contributions
  - $\mathcal{H}_2/\mathcal{H}_\infty$ norms convex in x — Dhingra, Colombino, Jovanović ECC '16
  - Leader selection — Dhingra, Colombino, Jovanović CDC '16
  - Robust control — Colombino, Dhingra, Jovanović, Smith CDC '16
  - Time-varying $x(t)$ — Dhingra, Colombino, Jovanović, Rantzer, Smith SCL '16

Combination drug therapy for HIV

Balance **performance** with **structure**



35 HIV mutants, 5 drugs

Budget	$\sum x_k \leq T$
Maximum dose	$x_k \leq T_k$
k requires j	$x_k \leq x_j$
Only k or j	$x_k + x_j \leq 1$

Model credit: Klein et al '12

Proximal operator

- Proximal operator

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

Proximal operator

- Proximal operator

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

- ℓ_1 norm – soft-thresholding

$$\mathbf{prox}_{\gamma\mu g}(v) \;=\; \begin{cases} v \,-\, \gamma\mu & v \;\geq\; \gamma\mu \\ 0 & |v| \;<\; \gamma\mu \\ v \,+\, \gamma\mu & v \;\leq\; -\gamma\mu \end{cases}$$

Parikh & Boyd, FnT Optimization '14

Proximal gradient method

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

Generalizes gradient descent

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

step-size α_k

Proximal gradient method

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

Generalizes gradient descent

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

step-size α_k

- convergence for f with Lipschitz cts gradient
- simple if $\mathbf{prox}_g$ easy to compute
- cannot be applied to $g(Tx)$
- acceleration with constraints (e.g., stability) challenging

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

Combination drug therapy – Budget

Optimize $f_2(x)$ or $f_\infty(x)$ subject to budget constraint

$$\begin{array}{rl} \text{minimize} & f(x) \\ \text{subject to} & \sum x_k \le 1 \\ & x_k \ge 0 \end{array}$$

Antibody	$x_{\mathcal{H}_2}$	$x_{\mathcal{H}_\infty}$
3BC176	0.5952	0.9875
PG16	0	0
45-46G54W	0.2484	0.0125
PGT128	0.1564	0
10-1074	0	0

Performance	$x_{\mathcal{H}_2}$	$x_{\mathcal{H}_\infty}$
$f_2(x)$	**0.6017**	1.1947
$f_\infty(x)$	0.1857	**0.1084**

Combination drug therapy – regularized

SELECT DRUGS:

$$x(\gamma) \;=\; \text{argmin} \quad f(x) \;+\; x^T x \;+\; \gamma \textstyle\sum_k w_k |x_k|$$

Increase γ until desired sparsity

POLISHING: design doses

$$\begin{array}{ll} \text{minimize} & f(x) \;+\; x^T x \\ \text{subject to} & \text{sp}(x) \;\in\; \text{sp}(x(\gamma)) \end{array}$$

Combination drug therapy – regularized

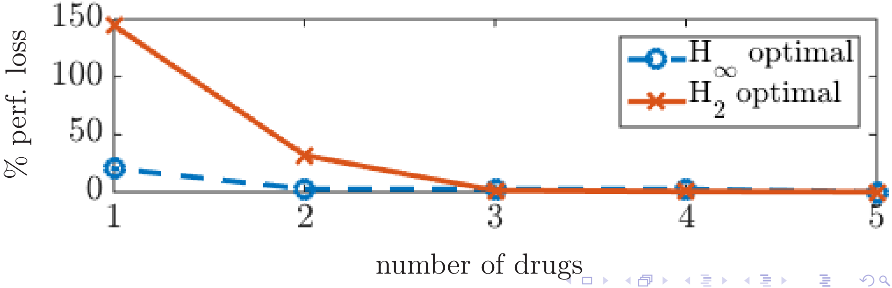
Select drugs:

$$x(\gamma) = \operatorname{argmin} \; \color{blue}{f(x)} + x^T x + \gamma \sum_k w_k |x_k|$$

Increase γ until desired sparsity

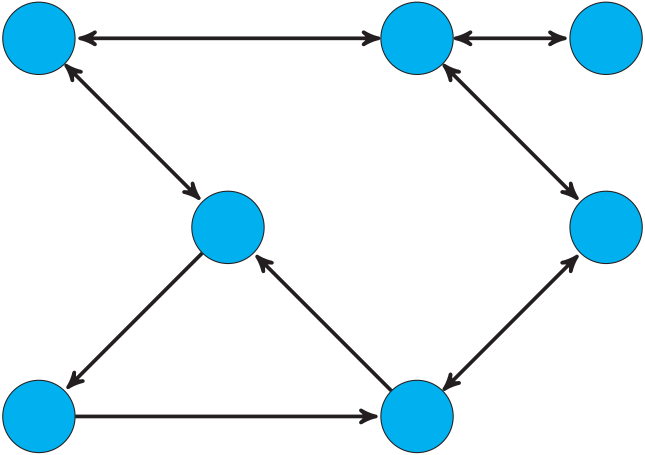
Polishing: design doses

$$\begin{aligned} \text{minimize} \quad & \color{blue}{f(x)} + x^T x \\ \text{subject to} \quad & \operatorname{sp}(x) \in \operatorname{sp}(x(\gamma)) \end{aligned}$$



Nonconvex regularized problems

Example: edge addition in directed consensus networks



- Distributed information exchange $\quad \dot{\psi} = -L_p\,\psi$
- If balanced, nodes approach average initial value, $\psi_i(t) \to \bar{\psi}$

Consensus networks

Dynamics

$$\dot{\psi} \;=\; -L_p\,\psi \,+\, d$$

Penalize deviation from average

$$\zeta \;=\; \left[\; I \,-\, (1/n)\mathbb{1}\mathbb{1}^T \;\right]\psi$$

Consensus networks

Dynamics

$$\dot{\psi} \;=\; -(L_p + L_c)\,\psi + d$$

Penalize deviation from average

$$\zeta \;=\; \begin{bmatrix} I - (1/n)\mathbb{1}\mathbb{1}^T \\ -\,R^{1/2}L_c \end{bmatrix}\psi$$

Add edges to network

- $F(w) = L_c$ – graph Laplacian of edges w
- $w = Tx$ parametrizes balanced graphs

Consensus networks

$$\underset{x}{\text{minimize}} \quad f_2(x) \; + \; \gamma \|Tx\|_1$$

Performance:

- $\mathcal{H}_2$ norm of deviations from average

Structure:

- Balanced L_c
- Minimize number of edges

Auxiliary variable

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

- benefit: **decouples** f and g

Augmented Lagrangian

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

Method of multipliers

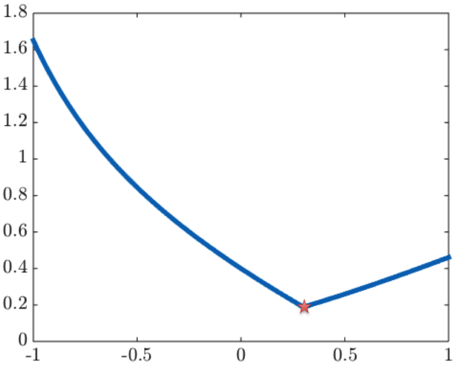
$$
\begin{aligned}
(x^{k+1}, z^{k+1}) &= \underset{x,z}{\operatorname{argmin}} \; \mathcal{L}_\mu(x,\, z;\, y^k) \\
y^{k+1} &= y^k + \frac{1}{\mu}\,(Tx^{k+1} - z^{k+1})
\end{aligned}
$$

Method of multipliers

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

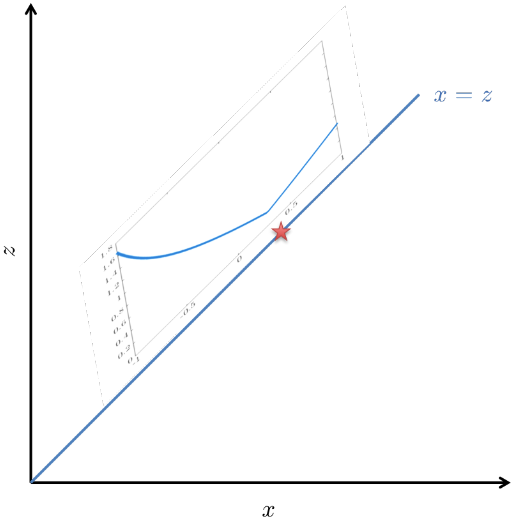
- guaranteed convergence to local minimum
- **challenge:** *joint* minimization over x and z

Method of multipliers cartoon



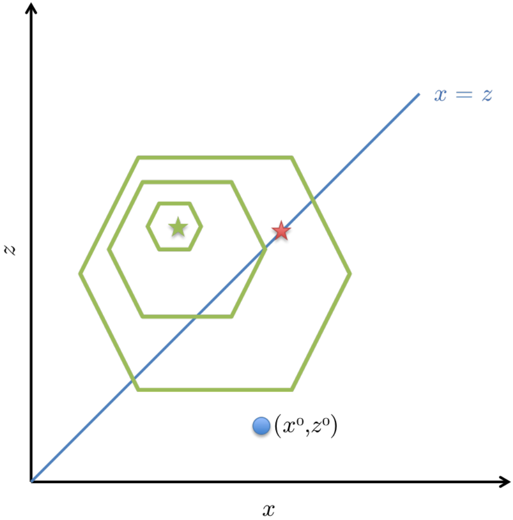
$f(x) + g(x)$

Method of multipliers cartoon



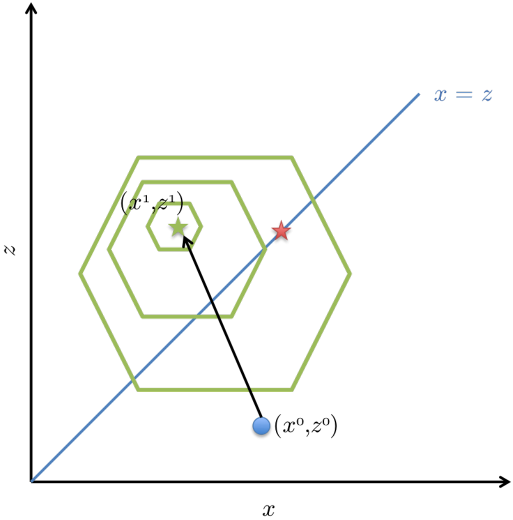
$f(x) + g(z)$

Method of multipliers cartoon



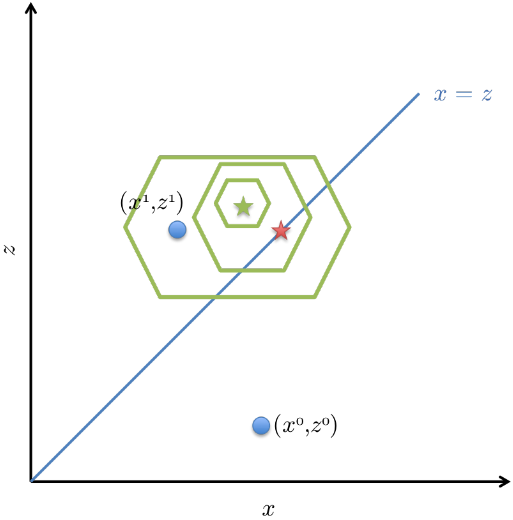
$\mathcal{L}_\mu(x, z; y^0)$

Method of multipliers cartoon



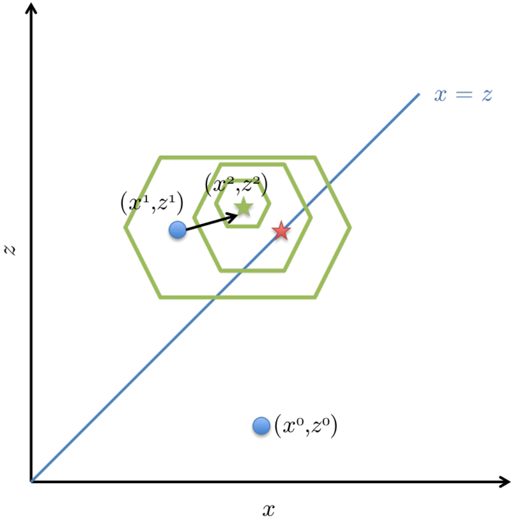
$\mathcal{L}_\mu(x, z; y^0)$

Method of multipliers cartoon



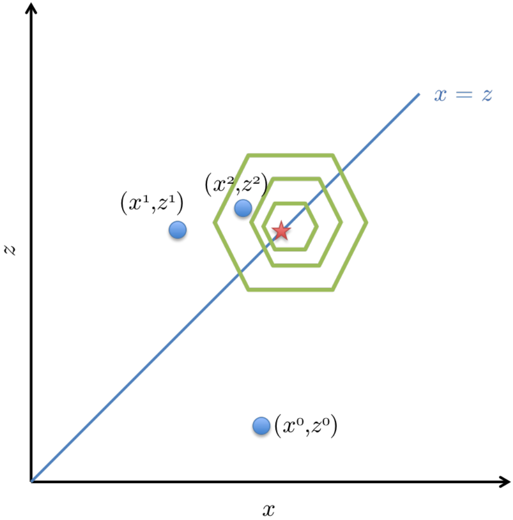
$\mathcal{L}_\mu(x, z; y^1)$

Method of multipliers cartoon



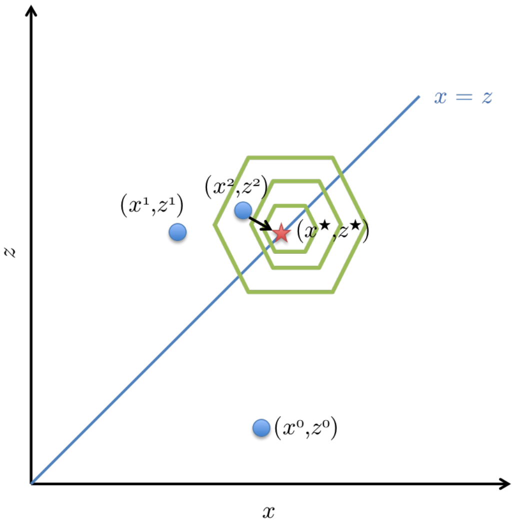
$\mathcal{L}_\mu(x, z; y^1)$

Method of multipliers cartoon



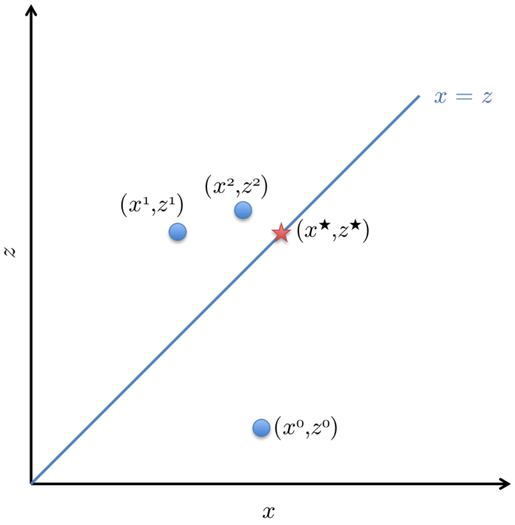
$\mathcal{L}_\mu(x, z; y^\star)$

Method of multipliers cartoon



$$\mathcal{L}_\mu(x, z; y^\star)$$

Method of multipliers cartoon



Alternating direction method of multipliers

$$
\begin{array}{rcll}
x^{k+1} & = & \underset{x}{\operatorname{argmin}} \; \mathcal{L}_\mu(x,\, z^k;\, y^k) & \text{differentiable} \\[2ex]
z^{k+1} & = & \underset{z}{\operatorname{argmin}} \; \mathcal{L}_\mu(x^{k+1},\, z;\, y^k) & \mathbf{prox}_{\gamma\mu g}(\cdot) \\[2ex]
y^{k+1} & = & y^k \; + \; \dfrac{1}{\mu}\,(Tx^{k+1} \; - \; z^{k+1}) &
\end{array}
$$

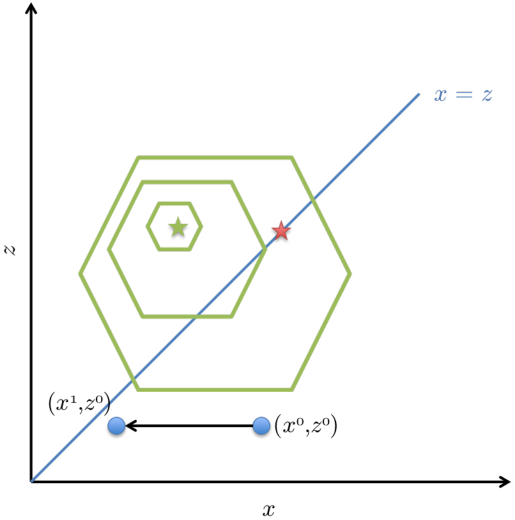
Alternating direction method of multipliers

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

- convenient for distributed implementation
- convergence speed influenced by μ
- **challenge:** convergence for nonconvex f

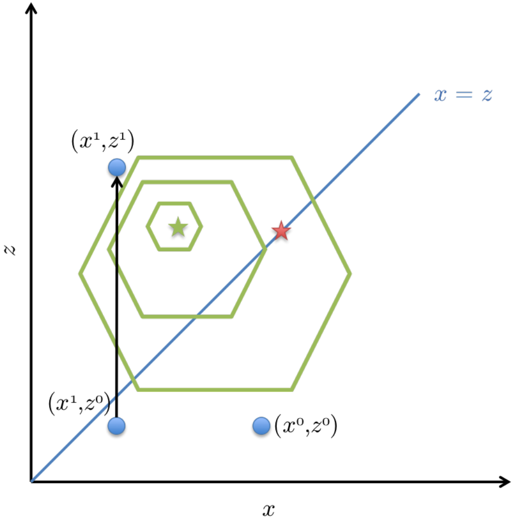
Hong, Luo, Razaviyayn, SIAM J. Optimiz. '16

Alternating direction method of multipliers cartoon



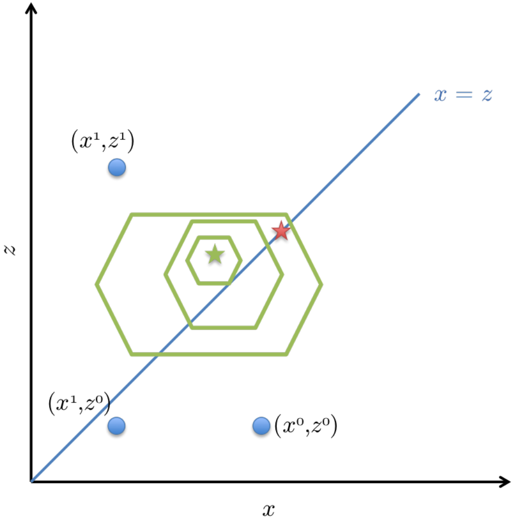
$\mathcal{L}_\mu(x, z; y^0)$

Alternating direction method of multipliers cartoon



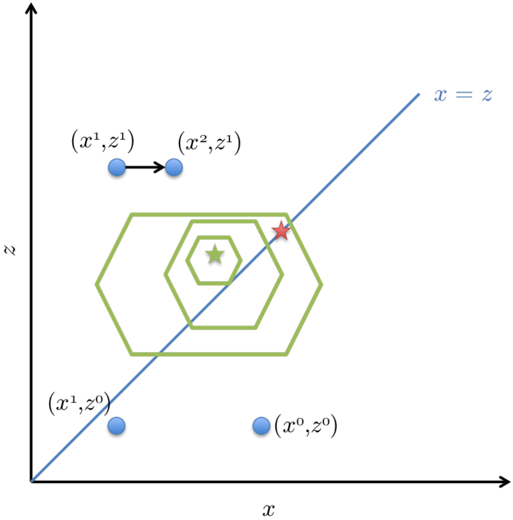
$\mathcal{L}_\mu(x, z; y^0)$

Alternating direction method of multipliers cartoon



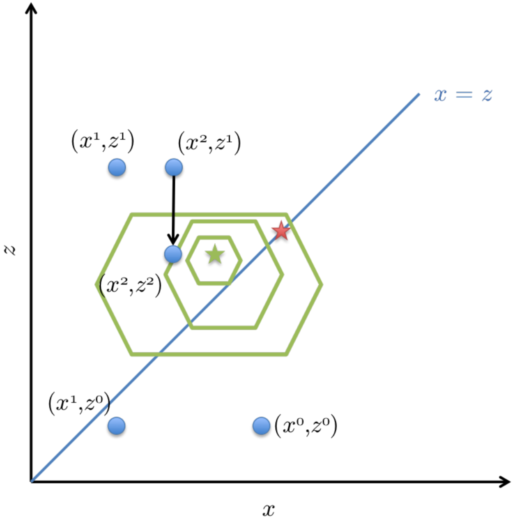
$\mathcal{L}_\mu(x, z; y^1)$

Alternating direction method of multipliers cartoon



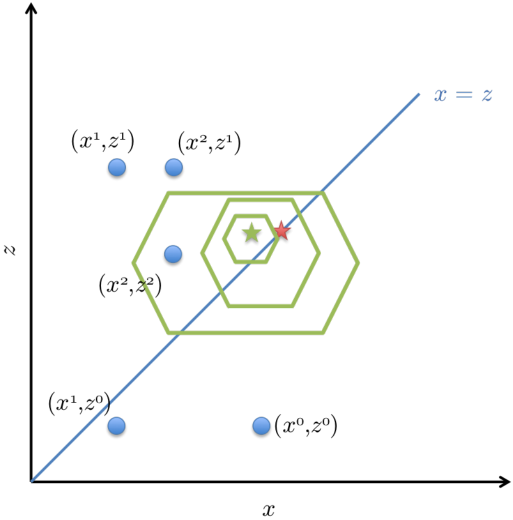
$\mathcal{L}_\mu(x, z; y^1)$

Alternating direction method of multipliers cartoon



$\mathcal{L}_\mu(x, z; y^1)$

Alternating direction method of multipliers cartoon



$\mathcal{L}_\mu(x, z; y^2)$

Alternating direction method of multipliers

$$
\begin{aligned}
x^{k+1} &= \operatorname*{argmin}_{x} \; \mathcal{L}_\mu(x,\, z^k;\, y^k) && \text{differentiable} \\
z^{k+1} &= \operatorname*{argmin}_{z} \; \mathcal{L}_\mu(x^{k+1},\, z;\, y^k) && \mathbf{prox}_{\gamma\mu g}(\cdot) \\
y^{k+1} &= y^k + \frac{1}{\mu}(Tx^{k+1} - z^{k+1})
\end{aligned}
$$

Proximal operator and Moreau envelope

- Proximal operator

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

Moreau envelope

$$M_{\gamma\mu g}(V) \;:=\; \inf_{x} \;\gamma\, g(x) \;+\; \frac{1}{2\mu}\, \|x \,-\, v\|^2$$

Proximal operator and Moreau envelope

- PROXIMAL OPERATOR

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

MOREAU ENVELOPE

$$M_{\gamma\mu g}(V) \;:=\; \inf_{x} \;\gamma\, g(x) \;+\; \frac{1}{2\mu}\,\|x \,-\, v\|^2$$

  - **continuously differentiable**
    even when g is not

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

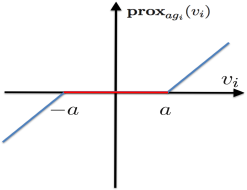
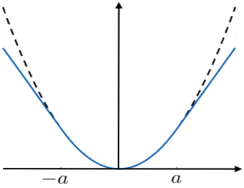
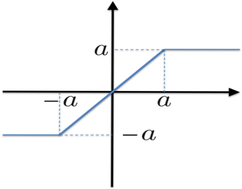
Parikh & Boyd, FnT in Optimization '14

Example

- Soft-thresholding – proximal operator 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
 	 	
	$a \; = \; \mu\gamma$	

Proximal augmented Lagrangian

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

MINIMIZE OVER z

$$z^\star \;=\; \mathbf{prox}_{\gamma\mu g}(Tx \;+\; \mu y)$$

Proximal augmented Lagrangian

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

MINIMIZE OVER z

$$z^\star \;=\; \mathbf{prox}_{\gamma\mu g}(Tx \;+\; \mu y)$$

EVALUATE $\mathcal{L}_\mu$ AT $z^\star$

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

continuously differentiable in x and y

Proximal augmented Lagrangian MM

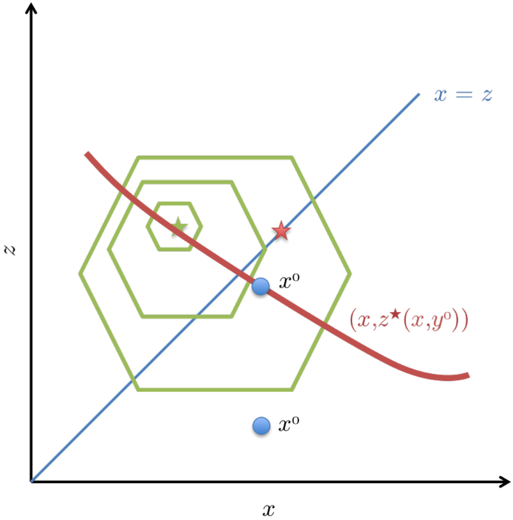
$$
\begin{aligned}
x^{k+1} &= \underset{x}{\operatorname{argmin}} \; \mathcal{L}_\mu(x;\, y^k) \\
y^{k+1} &= y^k \;+\; \frac{1}{\mu}\left(Tx^{k+1} \;-\; \mathbf{prox}_{\gamma\mu g}(Tx^{k+1} \;+\; \mu y^k)\right)
\end{aligned}
$$

Proximal augmented Lagrangian MM

$$
\begin{aligned}
x^{k+1} &= \underset{x}{\operatorname{argmin}} \ \mathcal{L}_\mu(x;\, y^k) \\
y^{k+1} &= y^k \ + \ \frac{1}{\mu}\left(Tx^{k+1} \ - \ \mathbf{prox}_{\gamma\mu g}(Tx^{k+1} \ + \ \mu y^k)\right)
\end{aligned}
$$

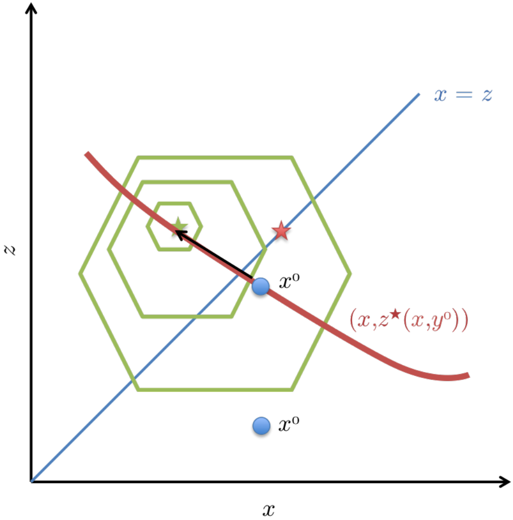
- nonconvex f: convergence to local minimum
- x-minimization step: differentiable problem
- adaptive μ update

Proximal augmented Lagrangian method of multipliers cartoon



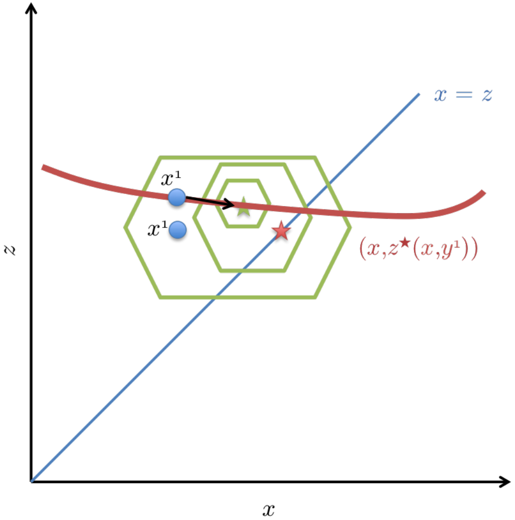
$\mathcal{L}_\mu(x, z; y^0)$

Proximal augmented Lagrangian method of multipliers cartoon



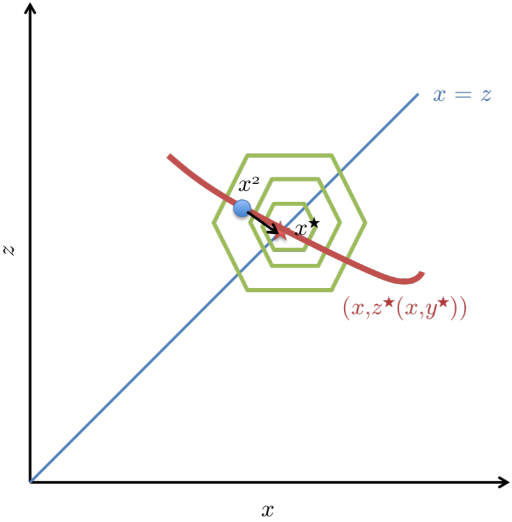
$\mathcal{L}_\mu(x, z; y^0)$

Proximal augmented Lagrangian method of multipliers cartoon



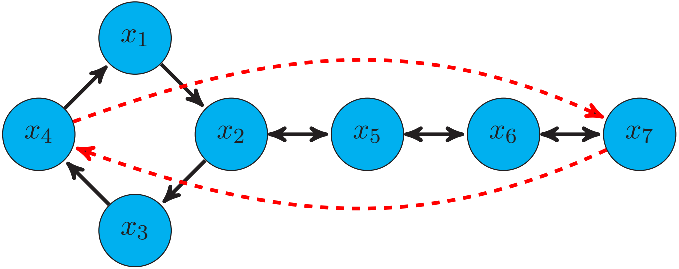
$\mathcal{L}_\mu(x, z; y^1)$

Proximal augmented Lagrangian method of multipliers cartoon



$\mathcal{L}_\mu(x, z; y^\star)$

Directed consensus network



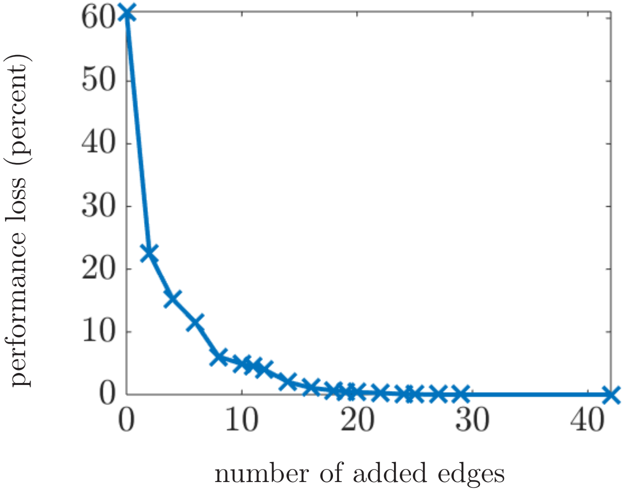
IDENTIFY EDGES

$$x(\gamma) \;=\; \underset{x}{\text{minimize}} \;\; f_2(x) \;+\; \gamma \|Tx\|_1$$

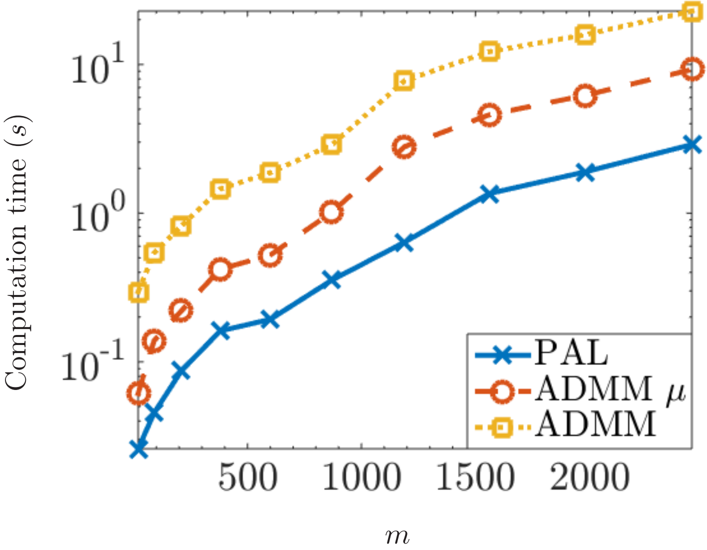
DESIGN EDGE WEIGHTS

$$\begin{array}{rll} x^\star(\gamma) \;=\; & \underset{x}{\text{minimize}} & f_2(x) \\ & \text{subject to} & \mathrm{sp}(Tx) \;\in\; \mathrm{sp}(Tx(\gamma)) \end{array}$$

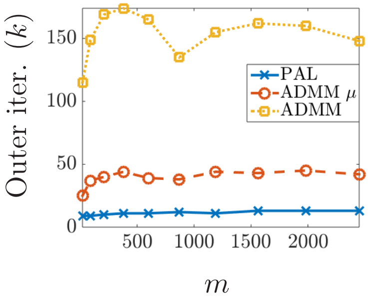
Directed consensus network



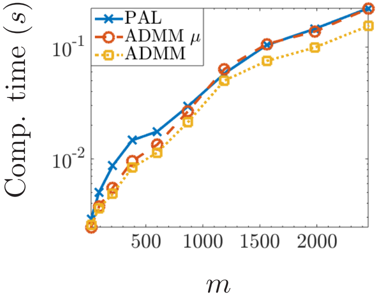
Comparison with ADMM



Comparison with ADMM



Outer iterations



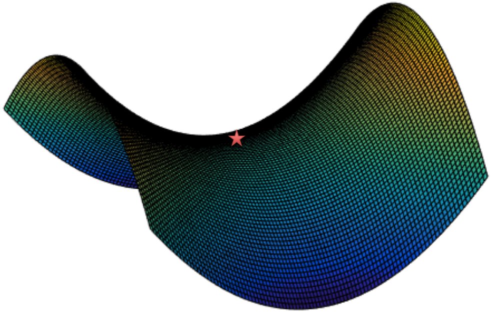
per outer iteration

Proximal augmented Lagrangian-based MM

- Guaranteed convergence
- Computational savings from reduced outer iterations

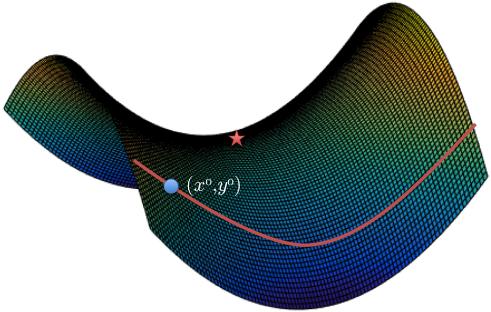
Convex regularized problems

Method of multipliers cartoon – primal-dual



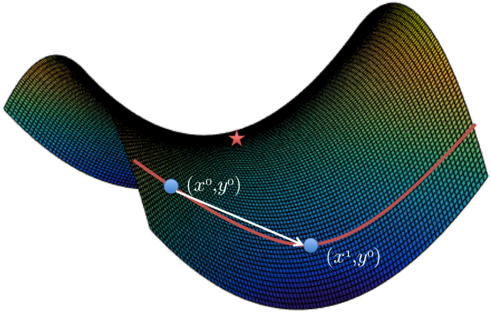
$\mathcal{L}_\mu(x; y)$

Method of multipliers cartoon – primal-dual



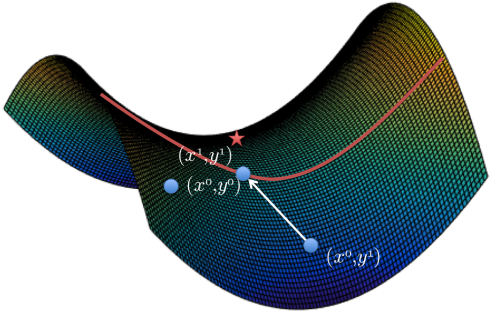
$$\mathcal{L}_\mu(x; y)$$

Method of multipliers cartoon – primal-dual



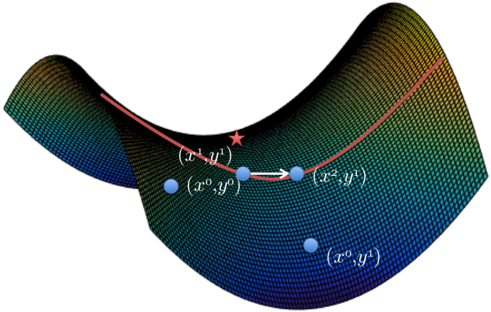
$$x^1 = \underset{x}{\operatorname{argmin}} \mathcal{L}_\mu(x; y^0)$$

Method of multipliers cartoon – primal-dual



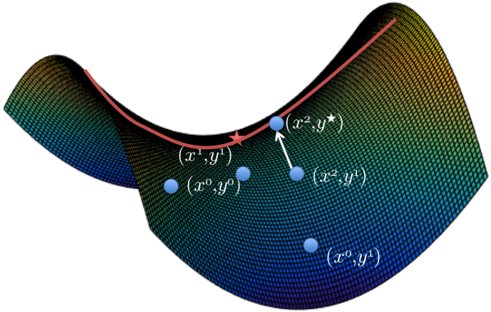
$$y^1 = y^0 + (1/\mu)\nabla_y \mathcal{L}_\mu(x^1; y^0)$$

Method of multipliers cartoon – primal-dual



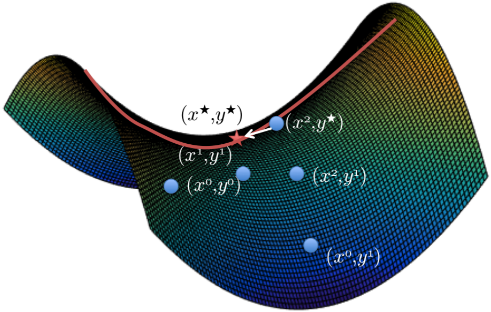
$$x^2 = \operatorname*{argmin}_{x} \mathcal{L}_\mu(x; y^1)$$

Method of multipliers cartoon – primal-dual



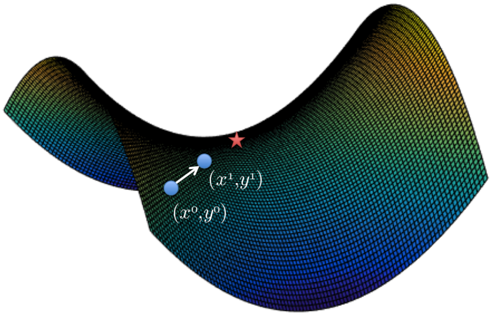
$$y^\star = y^1 + (1/\mu)\nabla_y \mathcal{L}_\mu(x^2; y^1)$$

Method of multipliers cartoon – primal-dual



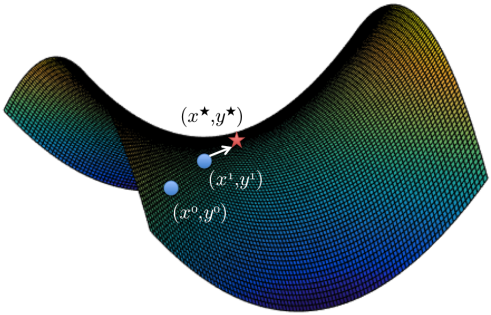
$$x^\star = \operatorname*{argmin}_x \mathcal{L}_\mu(x; y^\star)$$

Primal-descent dual-ascent cartoon



$$(x^1, y^1) = (x^0, y^0) - \alpha(\nabla_x \mathcal{L}_\mu(x^0; y^0), -\nabla_y \mathcal{L}_\mu(x^0; y^0))$$

Primal-descent dual-ascent cartoon



$$(x^\star, y^\star) = (x^1, y^1) - \alpha(\nabla_x \mathcal{L}_\mu(x^1; y^1), -\nabla_y \mathcal{L}_\mu(x^1; y^1))$$

Primal-descent dual-ascent

Arrow-Hurwicz-Uzawa type gradient flow

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

- continuous rhs even for non-differentiable g
- convenient for distributed implementation
- strictly convex – convergence
- strongly convex – conditions for exponential convergence

Arrow, Hurwicz, Uzawa, '59
Nedic & Ozdaglar, TAC '09
Wang & Elia, CDC '11
Feijer & Paganini, AUT '10
Cherukuri, Gharesifard, Cortés, SCL '15

Second-order updates

- First-order methods slow for high accuracy solutions
- Exploit second-order information

SECOND-ORDER METHODS FOR REGULARIZED PROBLEMS

- Proximal Newton — Lee, Sun, Saunders, '14
- Forward-backward envelope — Stella, Themelis, Patrinos, '16

Second-order updates

Assumptions:

- $T \in \mathbb{R}^{m \times n}$ full row rank
- g separable
- f twice cts diffble, strictly cvx

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

Second-order updates

ASSUMPTIONS:

- $T \in \mathbb{R}^{m \times n}$ full row rank
- g separable
- f twice cts diffble, strictly cvx

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

GENERALIZE DERIVATIVE OF $\mathbf{prox}_{\mu g}$ WITH P

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

n negative m positive eigenvalues

Dini derivatives

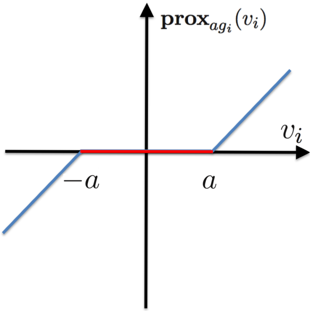
$P = \mathrm{diag}(p)$ and p_i is a Dini derivative $\mathbf{prox}_{\mu g}(v)$

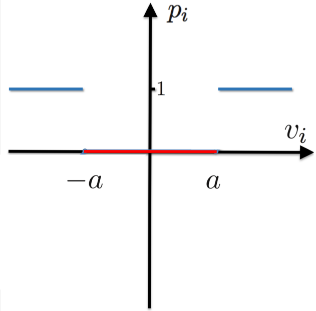
$$p_i \;=\; \lim_{\varepsilon \to 0^{\pm}} \frac{\mathbf{prox}_{\mu\gamma g}(v+\varepsilon) \;-\; \mathbf{prox}_{\mu\gamma g}(v)}{\varepsilon}$$

Dini derivatives

$P = \text{diag}(p)$ and p_i is a Dini derivative $\mathbf{prox}_{\mu g}(v)$

$$p_i \;=\; \lim_{\varepsilon \to 0^{\pm}} \frac{\mathbf{prox}_{\mu\gamma g}(v+\varepsilon) \;-\; \mathbf{prox}_{\mu\gamma g}(v)}{\varepsilon}$$





Soft-thresholding $\mathcal{S}_a(v)$

$$p \;\in\; \begin{cases} 0, & |v| \;<\; a \\ 1, & |v| \;>\; a \\ \{0,1\}, & |v| \;=\; a \end{cases}$$

Second-order updates

Continuous time – differential inclusion

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

- convergence to saddle point of $\mathcal{L}_\mu(x;y)$
- Lyapunov function $\|\nabla \mathcal{L}_\mu(x;y)\|^2$

Second-order updates

Continuous time – differential inclusion

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

- convergence to saddle point of $\mathcal{L}_\mu(x; y)$
- Lyapunov function $\|\nabla \mathcal{L}_\mu(x; y)\|^2$

Discrete time – algorithm

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

where $[\tilde{x}^T \; \tilde{y}^T]^T \in -(\nabla_P^2 \mathcal{L}_\mu)^{-1} \nabla \mathcal{L}_\mu(x; y)$

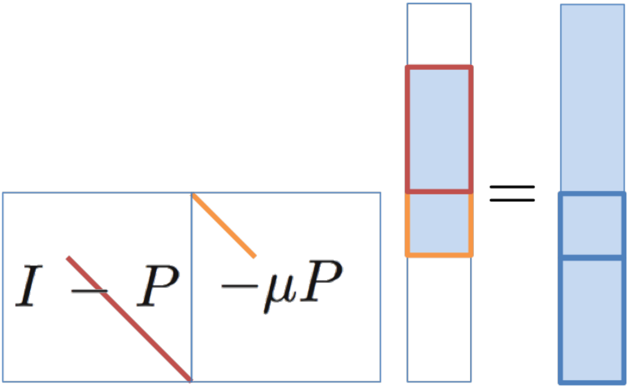
Efficient second-order updates for $g = \|\cdot\|_1$

$T = I$, $p_i \in \{0, 1\}$

$$\begin{bmatrix} \nabla^2 f & I \\ I - P & -\mu P \end{bmatrix} \begin{bmatrix} \tilde{x} \\ \tilde{y} \end{bmatrix} = b$$

Efficient second-order updates for $g = \|\cdot\|_1$

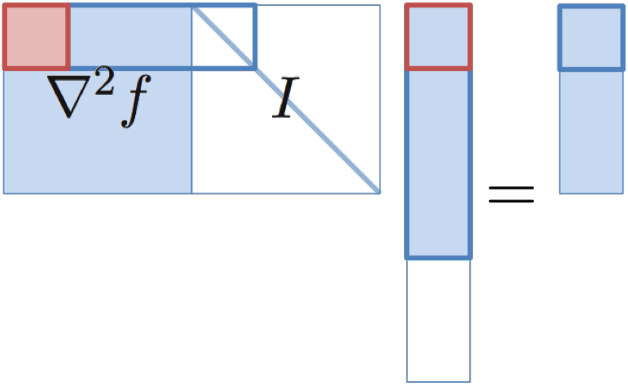
$T = I$, $p_i \in \{0, 1\}$



- Equality

Efficient second-order updates for $g = \|\cdot\|_1$

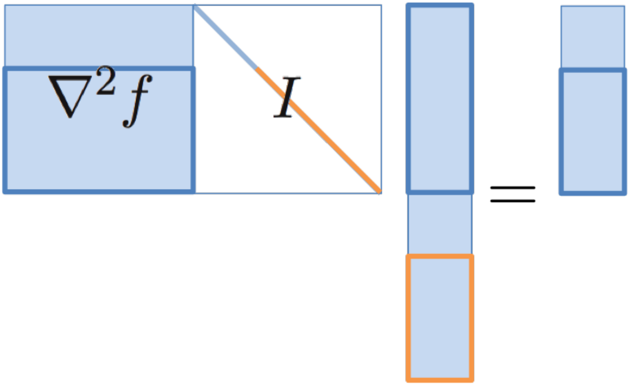
$T = I$, $p_i \in \{0, 1\}$



- Equality
- Limited matrix inversion (independent of μ)

Efficient second-order updates for $g = \|\cdot\|_1$

$T = I$, $p_i \in \{0, 1\}$



- Equality
- Limited matrix inversion (independent of μ)
- Matrix-vector multiplication

Second-order algorithm

Key question: how to assess progress?

- **Merit function:** primal-dual augmented Lagrangian

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

- y_e – Lagrange multiplier estimate

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

Second-order algorithm

Key question: how to assess progress?

- **Merit function:** primal-dual augmented Lagrangian

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

- y_e – Lagrange multiplier estimate

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

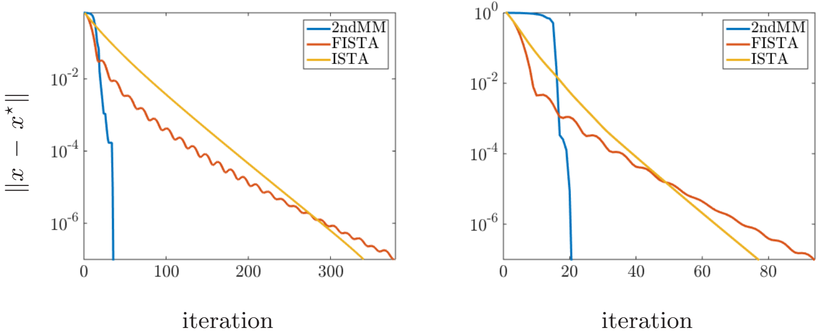
Adaptively decrease μ

Armand & Omheni, Optim. Method Softw. '15

LASSO

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

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

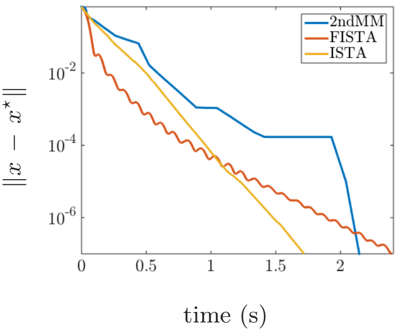


Note: require A^TA full rank

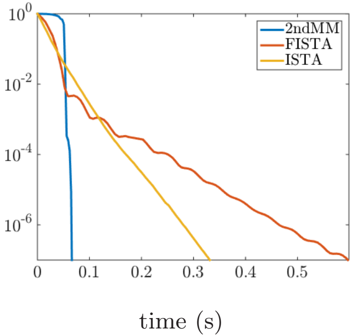
LASSO

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

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

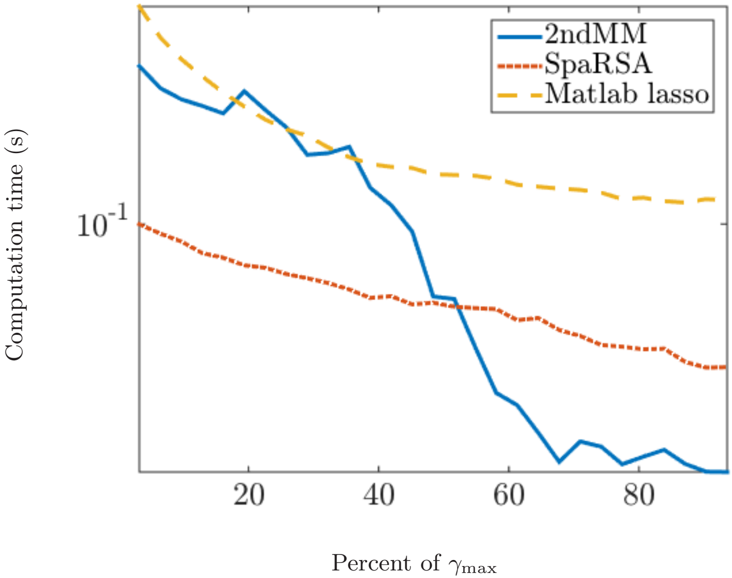


$\gamma = 0.85\gamma_{\max}$



Note: require A^TA full rank

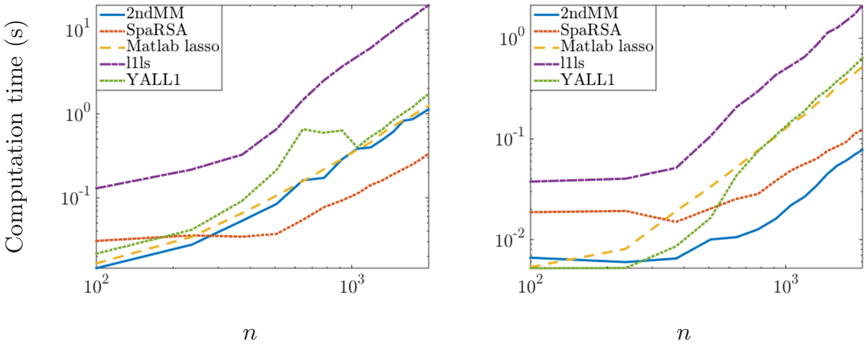
LASSO – dependence on γ



LASSO – dependence on size

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

$(\gamma = 0.15\gamma_{\max})$ $(\gamma = 0.85\gamma_{\max})$



$x \in \mathbb{R}^n$, $A \in \mathbb{R}^{2n\times n}$, A^TA full rank

Conclusions

Structured optimal control

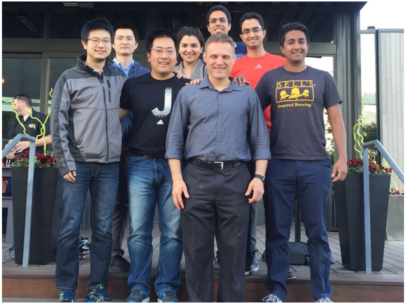
- Regularization to induce structure
- Convex problems
  - Actuator/sensor selection
  - Decentralized control of positive systems
  - Symmetric systems

Proximal augmented Lagrangian

- Differentiable – method of multipliers
- Arrow-Hurwicz-Uzawa updates
- Dini derivatives – second-order method

Acknowledgements

Acknowledgements



Acknowledgements

Acknowledgements