Exponential stability of primal-dual gradient flow dynamics based on proximal augmented Lagrangian

Dongsheng Ding

joint work with

Mihailo R. Jovanović

American Control Conference, Philadelphia, July 10-12, 2019

Nonsmooth composite minimization

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

T – select coordinates to impose structure

f – strongly convex; Lipschitz cts gradient

g – non-differentiable; convex

Examples

Optimization problem	$g(z)$
$\underset{x}{\text{minimize}} \quad f(x)$ $\text{subject to} \quad Tx = b$	$g(z) = \begin{cases} 0, & z = b \\ \infty, & \text{otherwise} \end{cases}$
$\underset{x}{\text{minimize}} \quad f(x)$ $\text{subject to} \quad Tx \leq b$	$g(z) = \begin{cases} 0, & z \leq b \\ \infty, & \text{otherwise} \end{cases}$
$\underset{x}{\text{minimize}} \quad f(x) + \gamma \, \lVert Tx \rVert_1$	$g(z) = \gamma \, \lVert z \rVert_1$

Proximal operator and Moreau envelope

Proximal operator

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

Moreau envelope

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

Parikh, Boyd, FnT in Optimization ’14

Proximal operator and Moreau envelope

Proximal operator

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

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 in v

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

Parikh, Boyd, FnT in Optimization '14

Augmented Lagrangian

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

$$\Downarrow$$

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

Augmented Lagrangian

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

Boyd, Parikh, Chu, Peleato, Eckstein, FnT in Optimization '11

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

Dhingra, Khong, Jovanović, IEEE TAC ’18

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

Dhingra, Khong, Jovanović, IEEE TAC '18

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

Minimizer of $\mathcal{L}_\mu(x, z; y)$ over z

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

Dhingra, Khong, Jovanović, IEEE TAC ’18

Proximal augmented Lagrangian

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

Minimizer of $\mathcal{L}_\mu(x,z;y)$ over z

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

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

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

Dhingra, Khong, Jovanović, IEEE TAC ’18

Proximal augmented Lagrangian

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

Minimizer of $\mathcal{L}_\mu(x, z; y)$ over z

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

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

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

continuously differentiable in x and y

Dhingra, Khong, Jovanović, IEEE TAC '18

Primal-dual gradient flow dynamics

Primal-descent dual-ascent

$$
\begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} = \begin{bmatrix} -\nabla_x \mathcal{L}_\mu(x;y) \\ \nabla_y \mathcal{L}_\mu(x;y) \end{bmatrix}
= \begin{bmatrix} -(\nabla f(x) + T^T \nabla M_{\mu g}(Tx + \mu y)) \\ \mu(\nabla M_{\mu g}(Tx + \mu y) - y) \end{bmatrix}
$$

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

- Lipschitz cts RHS
- $\bar{x} = x^\star, \ \bar{y} = y^\star$ – optimal solution

Dhingra, Khong, Jovanović, IEEE TAC '18

Related work

Global exponential stability

- Theory of IQCs
- Frequency-domain KYP Lemma
- Estimation of convergence rate

Dhingra, Khong, Jovanović, IEEE TAC ’18

Related work

Global exponential stability

- Problems with linear equality/inequality constraints
- Lyapunov-based characterization
- Estimation of convergence rate

Qu, Li, L-CSS '19

Related work

Global exponential stability

- Problems with linear equality/inequality constraints
- Lyapunov-based characterization
- Estimation of convergence rate

Qu, Li, L-CSS '19

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & T\,x \;=\; b \end{aligned}$$

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & T\,x \;\leq\; b \end{aligned}$$

Related work

Global exponential stability

- Problems with linear equality/inequality constraints
- Lyapunov-based characterization
- Estimation of convergence rate

Qu, Li, L-CSS '19

$$V(\tilde{w}) \;=\; \tilde{w}^T P \, \tilde{w}$$

$$P \;:=\; \begin{bmatrix} I & \alpha T^T \\ \alpha T & \beta I \end{bmatrix}$$

$$\tilde{w} := w - w^\star = \begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} x^\star \\ y^\star \end{bmatrix}$$

Primal-dual gradient flow dynamics

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

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

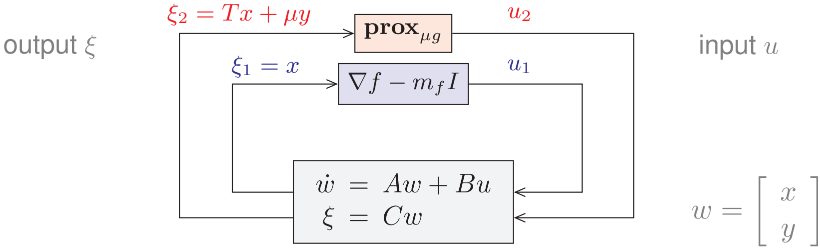
$$\Downarrow$$

$$\begin{bmatrix} \dot{x} \\ \dot{y} \end{bmatrix} = \begin{bmatrix} -(m_f I + \frac{1}{\mu} T^T T) & -T^T \\ T & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} -$$

$$\begin{bmatrix} I \\ 0 \end{bmatrix} (\nabla f(x) - m_f x) +$$

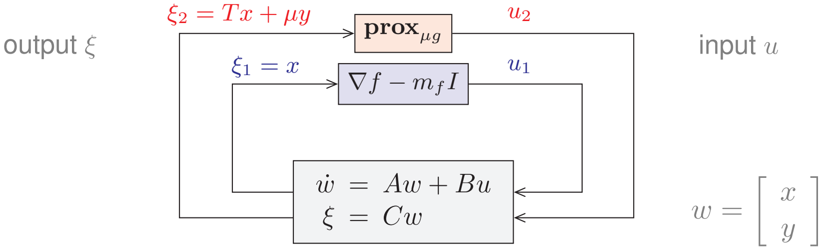
$$\begin{bmatrix} \frac{1}{\mu} T^T \\ -I \end{bmatrix} \mathbf{prox}_{\mu g}(Tx + \mu y)$$

Nonlinear feedback model



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

Nonlinear feedback model



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

LTI system

$$A \;=\; \begin{bmatrix} -(m_f I + \frac{1}{\mu} T^T T) & -T^T \\ T & 0 \end{bmatrix}$$

$$B \;=\; \begin{bmatrix} -I & \frac{1}{\mu} T^T \\ 0 & -I \end{bmatrix}, \quad C \;=\; \begin{bmatrix} I & 0 \\ T & \mu I \end{bmatrix}$$

Quadratic Lyapunov function

$$V(\tilde{w}) \; = \; \tilde{w}^T P \, \tilde{w}$$

$$P \; = \; \left[\begin{array}{cc} \alpha I & \beta T^T \\ \beta T & \gamma I \, + \, \zeta T T^T \end{array} \right]$$

$$\tilde{w} := w - w^\star = \left[\begin{array}{c} x \\ y \end{array} \right] - \left[\begin{array}{c} x^\star \\ y^\star \end{array} \right]$$

Quadratic Lyapunov function

$$V(\tilde{w}) \;=\; \tilde{w}^T P \, \tilde{w}$$

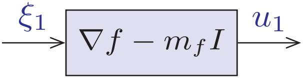
$$P \;=\; \alpha \left[\begin{array}{cc} I & \frac{1}{\mu} T^T \\ \frac{1}{\mu} T & (1 + \frac{m_f}{\mu}) I \,+\, \frac{1}{\mu^2} T T^T \end{array} \right] \;\succ\; 0$$

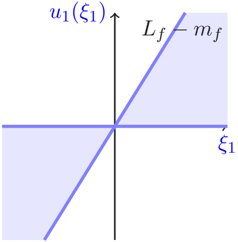
A – Hurwitz

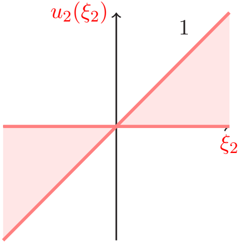
$$A^T P \,+\, P A \;=\; -2\alpha \left[\begin{array}{cc} m_f I & 0 \\ 0 & (1/\mu)\, T T^T \end{array} \right] \;\prec\; 0.$$

$$\tilde{w} := w - w^\star = \left[\begin{array}{c} x \\ y \end{array} \right] - \left[\begin{array}{c} x^\star \\ y^\star \end{array} \right]$$

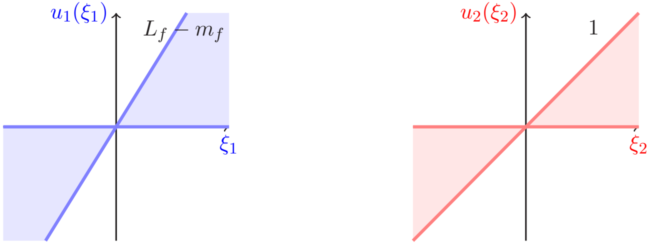
Sector nonlinearities







Sector nonlinearities



Pointwise quadratic constraints

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

Lessard, Recht, Packard, SIAM J. Opt. '16

Global exponential stability

$$\dot{V} = \begin{bmatrix} \tilde{w} \\ \tilde{u} \end{bmatrix}^T \begin{bmatrix} A^T P + PA & PB \\ B^T P & 0 \end{bmatrix} \begin{bmatrix} \tilde{w} \\ \tilde{u} \end{bmatrix}$$

- Quadratic constraint

$$\begin{bmatrix} \tilde{w} \\ \tilde{u} \end{bmatrix}^T \begin{bmatrix} 0 & C^T \Pi_0 \\ \Pi_0 C & -2\Lambda \end{bmatrix} \begin{bmatrix} \tilde{w} \\ \tilde{u} \end{bmatrix} \geq 0$$

Hu, Seiler, IEEE TAC '16

Global exponential stability

$$\dot{V} = \begin{bmatrix} \tilde{w} \\ \tilde{u} \end{bmatrix}^T \begin{bmatrix} A^T P + PA & PB \\ B^T P & 0 \end{bmatrix} \begin{bmatrix} \tilde{w} \\ \tilde{u} \end{bmatrix}$$

- Quadratic constraint

$$\begin{bmatrix} \tilde{w} \\ \tilde{u} \end{bmatrix}^T \begin{bmatrix} 0 & C^T \Pi_0 \\ \Pi_0 C & -2\Lambda \end{bmatrix} \begin{bmatrix} \tilde{w} \\ \tilde{u} \end{bmatrix} \geq 0$$

Exponential stability condition

$$\begin{bmatrix} -(A^T P + PA + 2\rho P) & -(PB + C^T \Pi_0) \\ -(PB + C^T \Pi_0)^T & 2\Lambda \end{bmatrix} \succeq 0$$

Hu, Seiler, IEEE TAC '16

Exponential convergence rate

$$\left[\begin{array}{cc} -(A^T P + PA + 2\rho P) & -(PB + C^T \Pi_0) \\ -(PB + C^T \Pi_0)^T & 2\Lambda \end{array}\right] \succeq 0$$

$$\left[\begin{array}{c} \tilde{w} \\ \tilde{u} \end{array}\right]^T \qquad \Downarrow \qquad \left[\begin{array}{c} \tilde{w} \\ \tilde{u} \end{array}\right]$$

$$\dot{V} \leq -2\rho V$$

Exponential decay

$$\|w(t) - w^\star\| \leq \sqrt{\kappa_P}\, \mathrm{e}^{-\rho t}\, \|w(0) - w^\star\|$$

Main result

Global exponential stability with rate $\rho > 0$

$$\|w(t) - w^\star\| \;\leq\; \sqrt{\kappa_P}\, \mathrm{e}^{-\rho t}\, \|w(0) - w^\star\|$$

$$\rho \;\geq\; \rho_0(\mu) \;:=\; \frac{1}{2}\,\frac{\sigma_{\min}(T)}{\mu + m_f + \frac{\sigma_{\max}(T)}{\mu}}$$

Main result

Global exponential stability with rate $\rho > 0$

$$\|w(t) - w^\star\| \leq \sqrt{\kappa_P}\, \mathrm{e}^{-\rho t}\, \|w(0) - w^\star\|$$

$$\rho \geq \rho_0(\mu) := \frac{1}{2} \frac{\sigma_{\min}(T)}{\mu + m_f + \frac{\sigma_{\max}(T)}{\mu}}$$

- $\mu \geq \max(\hat{L}, \hat{\mu})$
  - $\hat{L} = L_f - m_f > 0$
  - $\hat{\mu} \geq \sigma_{\max}(T), \; 2m_f \geq \frac{\sigma_{\max}^2(T)}{2\hat{\mu}}(1 + \frac{m_f}{\hat{\mu}}) + \frac{8\rho_0(\hat{\mu})^2}{\hat{\mu}} + 2\rho_0(\hat{\mu})$

Example

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & \tfrac{1}{2} x^T Q x + q^T x \\ \text{subject to} \quad & Tx \le b \end{aligned}$$

Example

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

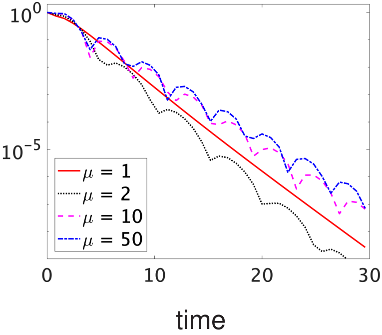
$f(x) = \frac{1}{2} x^T Q x + q^T x$

$g(z) = \{0,\ z \leq b$; ∞, otherwise$\}$

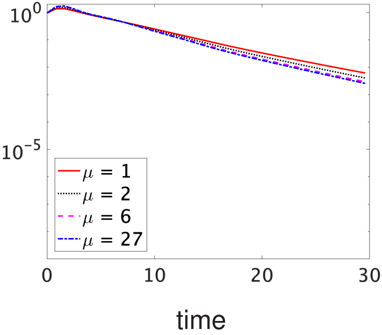
Exponential convergence

$$\frac{\|w(t) - w^\star\|}{\|w^\star\|}$$

$\kappa_Q \approx 1.2$



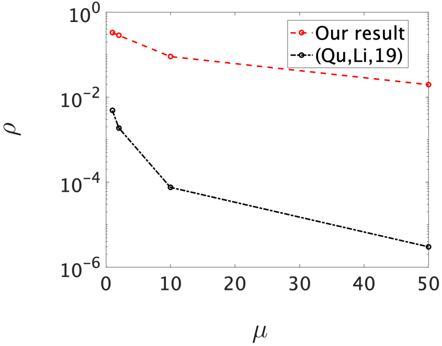
$\kappa_Q \approx 27.7$



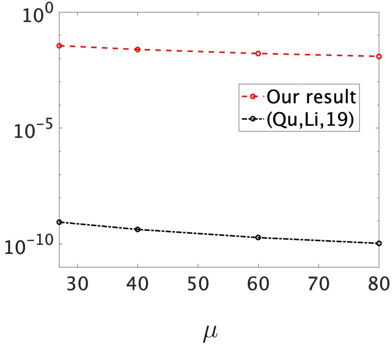
Convergence rate

$$\rho = \frac{1}{2}\frac{\sigma_{\min}(T)}{\mu + m_f + \frac{\sigma_{\max}(T)}{\mu}} \qquad \rho = \frac{\sigma_{\min}^4(T)}{80\sigma_{\max}^2(T) L_f \max\left(\frac{\mu\sigma_{\max}^2(T)}{m_f}, \frac{L_f}{m_f}\right)^2 \max\left(\frac{1}{\mu L_f}, \frac{L_f}{m_f}\right)^2}$$

$\kappa_Q \approx 1.2$



$\kappa_Q \approx 27.7$



Qu, Li, L-CSS '19

Summary

Primal-dual gradient flow dynamics

- Lyapunov-based convergence analysis
- Less conservative rates

Future work

- Optimal convergence rate
- Discretized algorithms