HARVARD
School of Engineering and Applied Sciences

ACC 2020

Exponential Stability of Primal-Dual Gradient Dynamics with Non-Strong Convexity

Xin Chen, Na Li

School of Engineering and Applied Sciences, Harvard University
chen_xin@g.harvard.edu

What is Primal-Dual Gradient Dynamics?

- Standard Convex Optimization

$$\min_{\mathbf{x}\in\mathbb{R}^n} f(\mathbf{x})$$
$$\text{s.t.} \quad \mathbf{g}(\mathbf{x}) \le \mathbf{0}$$
$$A\mathbf{x} = \mathbf{b}$$

- Lagrangian

$$L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\xi}) = f(\mathbf{x}) + \boldsymbol{\lambda}^T \mathbf{g}(\mathbf{x}) + \boldsymbol{\xi}^T (A\mathbf{x} - \mathbf{b})$$

- Max-min saddle point problem

$$\max_{\boldsymbol{\lambda}\ge\mathbf{0},\boldsymbol{\xi}} \min_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\xi})$$

Primal-Dual Gradient Dynamics

[Feijer and F. Paganini 2010]

$$\dot{\mathbf{x}} = -\nabla_{\mathbf{x}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\xi})$$
$$\dot{\boldsymbol{\lambda}} = \eta[\nabla_{\boldsymbol{\lambda}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\xi})]^+_{\boldsymbol{\lambda}}$$
$$\dot{\boldsymbol{\xi}} = \eta\nabla_{\boldsymbol{\xi}} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\xi})$$

Projection:

Stop decreasing when λ_j hits zero

$$\dot{\lambda}_j = \begin{cases} \eta\frac{\partial}{\partial\lambda_j} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\xi}) & \text{if } \lambda_j > 0 \\ \max(\eta\frac{\partial}{\partial\lambda_j} L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\xi}), 0) & \text{if } \lambda_j = 0 \end{cases}$$

PDGD

power system operation

wireless communication

resource allocation

Global Asymptotic Stability [Feijer and F. Paganini 2010] [Cherukuri, Mallada and Cortes 2016]

Linear Convergence for discrete-time counterpart [S. Du and W. Hu 2018]
[S. Du and J. Chen 2017]

Global Exponential Stability

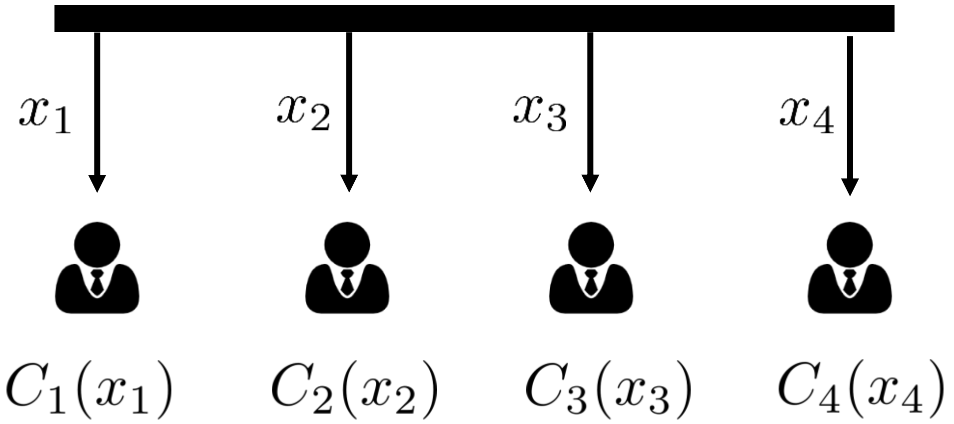
[Cortés and Niederländer 2017]	[Qu and Li 2018]	[Tang and Li 2019]
Equality constrained	**Linear inequality constrained**	**Convex inequality constrained**

Need the primal objective f(x) to be Strongly Convex.

Absent (Primal) Strong Convexity

Example 1. Resource Allocation

$$\min_{x_i \in \mathcal{X}_i} \sum_{i=1}^{n} C_i(x_i) \quad s.t. \quad \sum_{i=1}^{n} x_i = 1$$

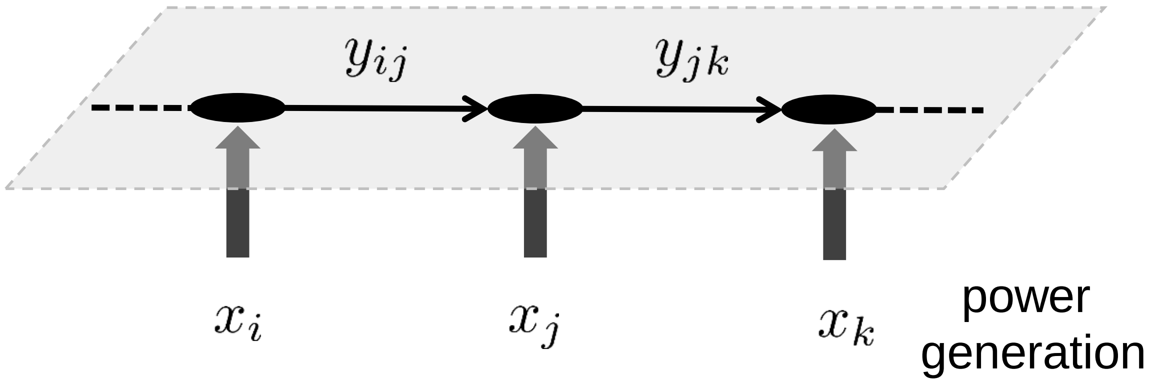


What if linear cost function for some agents

$$C_3(x_3) = a \cdot x_3 + b$$

Example 2. Optimal power network flow

$$\min_{\boldsymbol{x} \in \mathcal{X}, \boldsymbol{y} \in \mathcal{Y}} f(\boldsymbol{x}) \quad s.t. \quad A\boldsymbol{y} = \boldsymbol{x}$$



Obj. f(x) does not even contain variable y

Does PDGD still achieve global exponential stability?

Problem Formulation

Equality-constrained convex optimization problem

$$\min_{\boldsymbol{x}\in\mathbb{R}^n,\boldsymbol{y}\in\mathbb{R}^m} \quad f(\boldsymbol{x}) + g(\boldsymbol{y})$$
$$s.t. \qquad A\boldsymbol{x} + B\boldsymbol{y} = \boldsymbol{d}$$

f is strongly convex while g is only convex.

Lagrangian: $L(\boldsymbol{x},\boldsymbol{y},\boldsymbol{\lambda}) = f(\boldsymbol{x}) + g(\boldsymbol{y}) + \boldsymbol{\lambda}^\top (A\boldsymbol{x} + B\boldsymbol{y} - \boldsymbol{d})$

Primal-Dual Gradient Dynamics (PDGD)

$$\dot{\boldsymbol{x}} = -\eta_x \cdot \nabla_{\boldsymbol{x}} L(\boldsymbol{x},\boldsymbol{y},\boldsymbol{\lambda}) = -\eta_x \cdot \left(\nabla f(\boldsymbol{x}) + A^\top \boldsymbol{\lambda}\right)$$
$$\dot{\boldsymbol{y}} = -\eta_y \cdot \nabla_{\boldsymbol{y}} L(\boldsymbol{x},\boldsymbol{y},\boldsymbol{\lambda}) = -\eta_y \cdot \left(\nabla g(\boldsymbol{y}) + B^\top \boldsymbol{\lambda}\right)$$
$$\dot{\boldsymbol{\lambda}} = \quad \eta_\lambda \cdot \nabla_{\boldsymbol{\lambda}} L(\boldsymbol{x},\boldsymbol{y},\boldsymbol{\lambda}) = \quad \eta_\lambda \cdot (A\boldsymbol{x} + B\boldsymbol{y} - \boldsymbol{d})$$

Assumptions

$$\begin{array}{ll} \min_{\boldsymbol{x}\in\mathbb{R}^n, \boldsymbol{y}\in\mathbb{R}^m} & f(\boldsymbol{x}) + g(\boldsymbol{y}) \\ s.t. & A\boldsymbol{x} + B\boldsymbol{y} = \boldsymbol{d} \end{array}$$

Assumption 1: f is μ-strongly convex and ℓ-smooth, g is convex and ρ-smooth, i.e.,

$$\mu||\boldsymbol{x}_1 - \boldsymbol{x}_2||^2 \leq \langle \nabla f(\boldsymbol{x}_1) - \nabla f(\boldsymbol{x}_2), \boldsymbol{x}_1 - \boldsymbol{x}_2 \rangle \leq \ell||\boldsymbol{x}_1 - \boldsymbol{x}_2||^2$$

$$0 \leq \langle \nabla g(\boldsymbol{y}_1) - \nabla g(\boldsymbol{y}_2), \boldsymbol{y}_1 - \boldsymbol{y}_2 \rangle \leq \rho||\boldsymbol{y}_1 - \boldsymbol{y}_2||^2$$

Assumption 2: The convex problem has a finite optimum.

Assumption 3: A is full row rank and $\kappa_1 I \preceq AA^\top \preceq \kappa_2 I$ for some $0 < \kappa_1 \leq \kappa_2$.

(Standard assumption for exponential stability. Matrix A is the bridge between $\boldsymbol{x}$ and $\boldsymbol{\lambda}$)

In this setting, is the PDGD globally exponentially stable?

Intuitions

$\boldsymbol{z}^* := \left[\boldsymbol{x}^{*\top}, \boldsymbol{y}^{*\top}, \boldsymbol{\lambda}^{*\top}\right]^\top$: an equilibrium point

$\Delta\boldsymbol{x} = \boldsymbol{x} - \boldsymbol{x}^*$

$\Delta\boldsymbol{y} = \boldsymbol{y} - \boldsymbol{y}^*$

$\Delta\boldsymbol{\lambda} = \boldsymbol{\lambda} - \boldsymbol{\lambda}^*$

❑ The prima-dual gradient dynamics:

$$\dot{\boldsymbol{x}} = -\nabla f(\boldsymbol{x}) - A^\top \boldsymbol{\lambda} = -(\nabla f(\boldsymbol{x}) - \nabla f(\boldsymbol{x}^*)) - A^\top \Delta\boldsymbol{\lambda}$$

Special Case $B = 0$

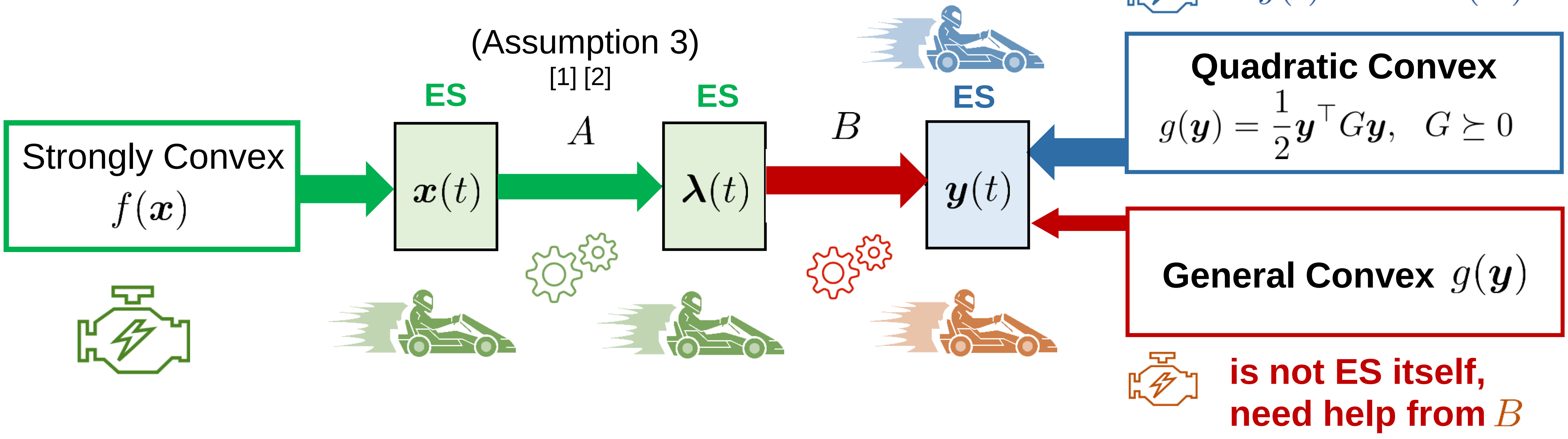
$$\boxed{\dot{\boldsymbol{y}} = -\nabla g(\boldsymbol{y}) - B^\top \boldsymbol{\lambda} = -(\nabla g(\boldsymbol{y}) - \nabla g(\boldsymbol{y}^*))} - B^\top \Delta\boldsymbol{\lambda}$$

$$\dot{\boldsymbol{\lambda}} = A\boldsymbol{x} + B\boldsymbol{y} - \boldsymbol{d} = A\Delta\boldsymbol{x} + B\Delta\boldsymbol{y}$$

$\Rightarrow$ **linear system** $\dot{\Delta\boldsymbol{y}} = G\Delta\boldsymbol{y}$

$\Downarrow$

$\Delta y(t) \xrightarrow{ES} ker(G)$



[1] [J. Cortés, S. K. Niederländer, 2018] [2] [G. Qu, N. Li, 2018]

Quadratic Case

Consider a quadratic $g(\boldsymbol{y})$ given by $$g(\boldsymbol{y}) = \frac{1}{2}\boldsymbol{y}^\top G\boldsymbol{y} + \boldsymbol{g}^\top \boldsymbol{y} + g_0$$

$$\begin{aligned} \dot{\boldsymbol{x}} &= -\eta_x \cdot \left(\nabla f(\boldsymbol{x}) + A^\top \boldsymbol{\lambda}\right) \\ \dot{\boldsymbol{y}} &= -\eta_y \cdot \left(\nabla g(\boldsymbol{y}) + B^\top \boldsymbol{\lambda}\right) \\ \dot{\boldsymbol{\lambda}} &= \quad \eta_\lambda \cdot (A\boldsymbol{x} + B\boldsymbol{y} - \boldsymbol{d}) \end{aligned}$$

with $g_0 \in \mathbb{R}$, $\boldsymbol{g} \in \mathbb{R}^m$, $G \in \mathbb{S}^{m\times m}$ and $0 \preceq G \preceq \rho I$.

$$\boldsymbol{z} := [\boldsymbol{x}; \boldsymbol{y}; \boldsymbol{\lambda}]$$

Proposition 1. (informal) The equilibrium point set of the PDGD is given by

$$\Psi := \{\hat{\boldsymbol{z}} | \ \hat{\boldsymbol{x}} = \boldsymbol{x}^*, \hat{\boldsymbol{\lambda}} = \boldsymbol{\lambda}^*, \underbrace{B\hat{\boldsymbol{y}} = B\boldsymbol{y}^*, G\hat{\boldsymbol{y}} = G\boldsymbol{y}^*}_{y^* \text{ is not unique } \hat{\boldsymbol{y}} - \boldsymbol{y}^* \in ker(B) \cap ker(G)} \}$$

Theorem 1. Under assumption 1, 2, 3 and the condition that $g(\boldsymbol{y})$ is quadratic, then the PDGD is globally exponentially stable, i.e., there exists $a_z \geq 0, \tau > 0$ such that

$$\mathrm{dist}(\boldsymbol{z}(t), \Psi) \leq a_z \cdot e^{-\tau t}$$

Note 1. PDGD is still nonlinear due to general f(x).

Note 2. The linear case with $G = 0$ is prohibited by Assumption 2 (finite optimum).

General Case

General convex smooth $g(\boldsymbol{y})$

$g(\boldsymbol{y})$ *may not ensure exponential stability itself and needs help from B.*

Assumption 4 (Augmented Strong Convexity): For any $\boldsymbol{y}_1, \boldsymbol{y}_2$, there exists $\gamma > 0$ such that

$$\underbrace{(\boldsymbol{y}_1 - \boldsymbol{y}_2)^\top B^\top B(\boldsymbol{y}_1 - \boldsymbol{y}_2) + \langle \nabla g(\boldsymbol{y}_1) - \nabla g(\boldsymbol{y}_2), \boldsymbol{y}_1 - \boldsymbol{y}_2 \rangle}_{B \text{ and } g(\boldsymbol{y}) \text{ together mimic a strong convexity condition.}} \geq \gamma \cdot ||\boldsymbol{y}_1 - \boldsymbol{y}_2||^2$$

Augmented $\hat{g}(\boldsymbol{y}) := \boldsymbol{y}^\top B^\top B \boldsymbol{y} + g(\boldsymbol{y})$ is strongly convex.

Proposition 2. (informal) The equilibrium point $\boldsymbol{z}^*$ of the PDGD is **unique** under Assumption 1 - 4.

Theorem 2. Under Assumption 1, 2, 3, 4, the PDGD is globally exponentially stable, i.e., there exists $c_z \geq 0$ and $\tau > 0$ such that

$$||\boldsymbol{z}(t) - \boldsymbol{z}^*|| \leq c_z \cdot e^{-\tau t}$$

Proof Sketch

Define $\boldsymbol{z} := [\boldsymbol{x}^\top, \boldsymbol{y}^\top, \boldsymbol{\lambda}^\top]^\top$, $\boldsymbol{z}^* := [\boldsymbol{x}^{*\top}, \boldsymbol{y}^{*\top}, \boldsymbol{\lambda}^{*\top}]^\top$.

❖ ***Challenge***: there is no strong convexity on primal variable **y** in the Lagrangian.

❖ ***Key Idea***: (use matrix B to make up strong convexity deficit.)

Find a Lyapunov function $V(\boldsymbol{z}) \geq 0$ such that $\frac{dV(\boldsymbol{z})}{dt} \leq -\tau V(\boldsymbol{z})$ for constant $\tau > 0$.

Quadratic Case

Design $V_1(\boldsymbol{z}) = (\boldsymbol{z} - \boldsymbol{z}^*)^\top P_1 (\boldsymbol{z} - \boldsymbol{z}^*)$

where $P_1 = \begin{bmatrix} \frac{\alpha}{\eta_x} I & 0 & \boxed{\frac{1}{\eta_\lambda} A^\top} \\ 0 & \frac{\alpha}{\eta_y} U^\top U & \boxed{-\frac{\beta}{\eta_\lambda} B^\top} \\ \boxed{\frac{1}{\eta_\lambda} A} & \boxed{-\frac{\beta}{\eta_\lambda} B} & \frac{\alpha}{\eta_\lambda} I \end{bmatrix} \succeq 0$

General Case

Design $V_2(\boldsymbol{z}) = (\boldsymbol{z} - \boldsymbol{z}^*)^\top P_2 (\boldsymbol{z} - \boldsymbol{z}^*)$

where $P_2 = \begin{bmatrix} \frac{\alpha}{\eta_x} I & 0 & \frac{1}{\eta_\lambda} A^\top \\ 0 & \frac{\alpha}{\eta_y} I & -\frac{\beta}{\eta_\lambda} B^\top \\ \frac{1}{\eta_\lambda} A & -\frac{\beta}{\eta_\lambda} B & \frac{\alpha}{\eta_\lambda} I \end{bmatrix} \succeq 0$

- U is orthonormal in column and $ker(U) = ker(B) \cap ker(G)$.
- α is a sufficiently large parameter to make diagonal dominant.

Simulation (Quadratic Case)

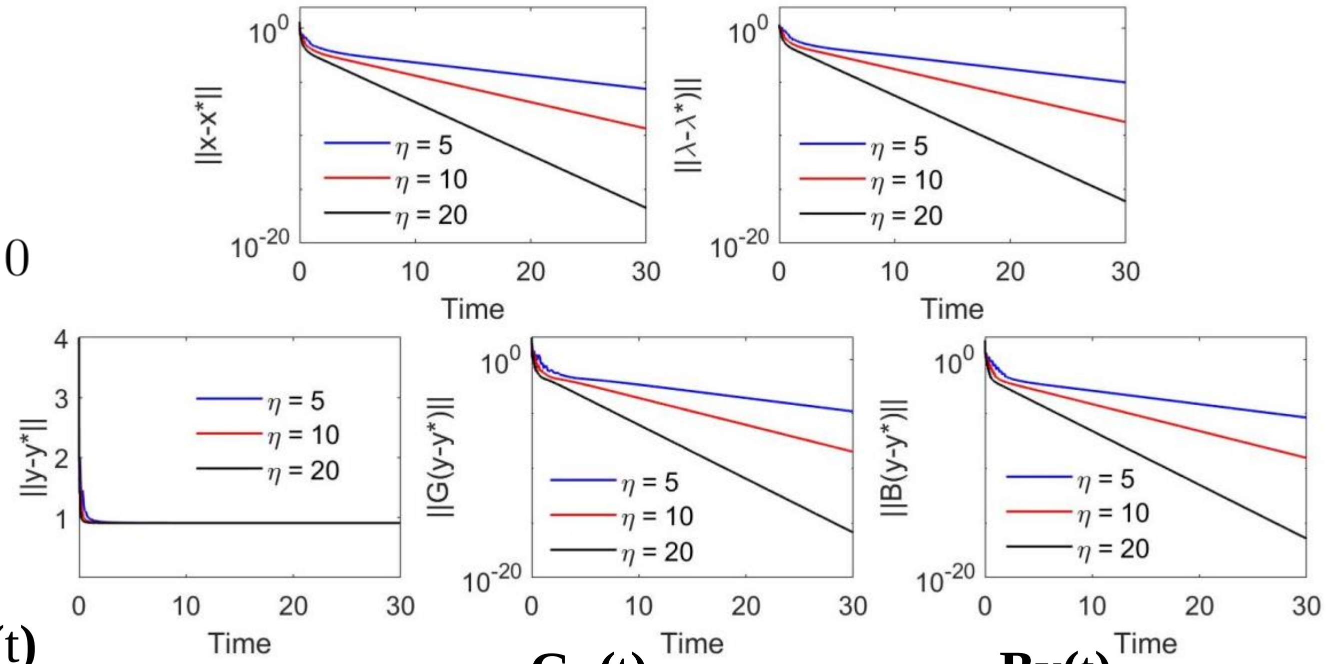
$$\boxed{\begin{aligned} \min_{\boldsymbol{x}\in\mathbb{R}^n, \boldsymbol{y}\in\mathbb{R}^m} \quad & f(\boldsymbol{x}) + g(\boldsymbol{y}) \\ s.t. \quad & A\boldsymbol{x} + B\boldsymbol{y} = \boldsymbol{d} \end{aligned}}$$

- $g(\boldsymbol{y}) = \frac{1}{2}\boldsymbol{y}^\top G\boldsymbol{y}, \quad G = \mathrm{diag}(0, G_0^\top G_0) \succeq 0$
- $f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^\top F\boldsymbol{x}, \quad F \succ 0$
- $B := [\mathbf{0}, B_0]$

$$\boldsymbol{e}_1 = [1, 0, \cdots, 0]^\top$$

$$\boldsymbol{e}_1 \in ker(G) \cap ker(B)$$

$\Delta\boldsymbol{x}(t) \to 0$ $\Delta\boldsymbol{\lambda}(t) \to 0$



y(t) **Gy**(t) **By**(t)

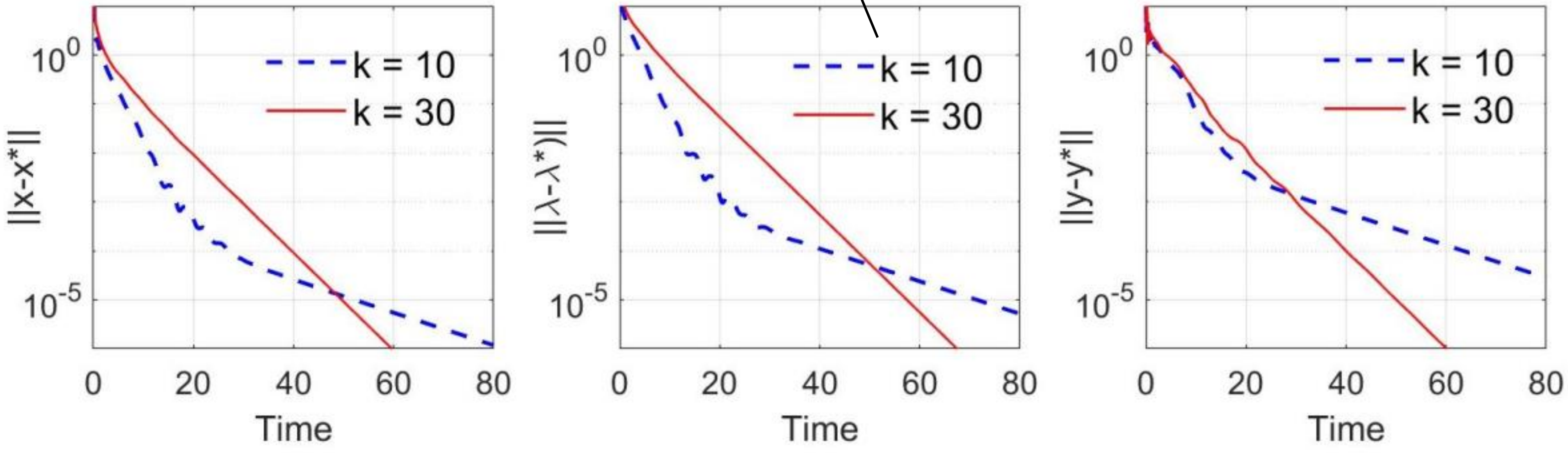
Fig. 1. Convergence results of PDGD with different time constants when $g(\boldsymbol{y})$ is a quadratic function.

$$\Delta\boldsymbol{y}(t) = \boldsymbol{y}(t) - \boldsymbol{y}^* \xrightarrow{\text{ES}} ker(B) \cap ker(G)$$

Simulation (General Case)

$$\min_{\boldsymbol{x}\in\mathbb{R}^n,\boldsymbol{y}\in\mathbb{R}^m} \quad f(\boldsymbol{x})+g(\boldsymbol{y})$$
$$s.t. \quad A\boldsymbol{x}+B\boldsymbol{y}=\boldsymbol{d}$$

- $f(\boldsymbol{x})=\frac{1}{2}\boldsymbol{x}^\top F\boldsymbol{x}, \quad F\succ 0$
- $g(\boldsymbol{y})=\sum_{i=1}^m y_i^4$
- $B\in\mathbb{R}^{k\times 20}$



k = 10	B is not full column rank.	Assum. 4 is **not** satisfied.	**Locally** exponentially stable.
k = 30	B is full column rank.	Assum. 4 is **satisfied.**	**Globally** exponentially stable.

Conclusion

$$\min_{\boldsymbol{x}\in\mathbb{R}^n,\boldsymbol{y}\in\mathbb{R}^m} \quad f(\boldsymbol{x})+g(\boldsymbol{y})$$
$$s.t. \quad A\boldsymbol{x}+B\boldsymbol{y}=\boldsymbol{d}$$

f is smooth and strongly convex
g is smooth and convex
A is of full row rank

if ($g(\boldsymbol{y})$ **is quadratic**) **or** ($g(\boldsymbol{y})$ **and B satisfy augmented strong convexity**)

The PDGD can achieve global exponential stability;

end

Future work

Extend the analysis to convex optimization with (non)linear inequality constraints.

Thanks!

Some References

[1] A. Cherukuri, E. Mallada, and J. Cortés, “Asymptotic convergence of constrained primal-dual dynamics,” Systems & Control Letters, vol. 87,pp. 10-15, 2016.
[2] J. Cortés and S. K. Niederländer, “Distributed coordination for non-smooth convex optimization via saddle-point dynamics,” Journal of Nonlinear Science, pp. 1-26, 2018.
[3] G. Qu and N. Li, “On the exponential stability of primal-dual gradient dynamics,” *arXiv preprint*, arXiv:1803.01825, 2018.
[4] S. S. Du, W. Hu, “Linear convergence of the primal-dual gradient method for convex-concave saddle point problems without strong convexity,” arXiv preprint, arXiv:1802.01504, 2018.
[5] N. K. Dhingra, S. Z. Khong, and M. R. Jovanovic, “The proximal augmented Lagrangian method for nonsmooth composite optimization,” IEEE Transactions on Automatic Control, 2018.
[6] X. Chen and N. Li, “Exponential stability of primal-dual gradient dynamics with non-strong convexity,” arXiv preprint, arXiv:1905.00298, 2020.