A Family of ODE-solver based Fast ADMM with Applications to Sparse Regression

Hao Luo (罗浩)

National Center for Applied Mathematics in Chongqing
Chongqing Normal University
Email: luohao@cqnu.edu.cn

Workshop on Scientific Computing and Data Science

Department of Mathematics Chinese University of Hong Kong

July 11–12, 2024

A joint work with Dr. Zi-Hang Zhang from PKU

Outline

Introduction

PDHG and CP

Primal-Dual Flow Dynamics

Accelerated Primal-Dual Flow

Time Discretization

Numerical Results

Problem setting

- Equality constrained separable minimization

$$\boxed{\min_{u\in\mathcal{U},v\in\mathcal{V}} F(u,v) := f(u)+g(v) \quad \text{s.t. } Au+Bv=b} \qquad (1)$$

Assumptions:

- $\mathcal{U}$, $\mathcal{V}$, Λ: Hilbert spaces with inner product $\langle\cdot,\cdot\rangle$ [1]
- $A(B):\mathcal{U}(\mathcal{V})\to\Lambda$: bounded linear operators, $b\in\Lambda$
- $f(g):\mathcal{U}(\mathcal{V})\to(-\infty,+\infty]$: CCP [2] with constants $\mu_f(\mu_g)\geq 0$
- Consistent condition: $b\in A\,\mathbf{dom}\,f+B\,\mathbf{dom}\,g$

- Composite convex minimization $\left(b=0,\ B=-I\right)$

$$\boxed{\min_{u\in\mathcal{U}} P(u) := f(u)+g(Au)} \qquad (2)$$

[1] When no confusion arises, we use the same bracket $\langle\cdot,\cdot\rangle$ for the inner products on $\mathcal{U}$, $\mathcal{V}$ and Λ.

[2] CCP means closed, convex and proper.

Preliminary

- Introduce $\boldsymbol{x} = (u, v)$, $\boldsymbol{A} = (A, B)$ and restate the single block form

$$\min_{\boldsymbol{x}\in\mathcal{U}\times\mathcal{V}} F(\boldsymbol{x}) \quad \text{s.t. } \boldsymbol{A}\boldsymbol{x} = b \tag{3}$$

- Define the Lagrangian

$$\mathcal{L}(\mathbf{x}, \lambda) := F(\mathbf{x}) + \langle \lambda, \boldsymbol{A}\boldsymbol{x} - b \rangle\,, \quad (\boldsymbol{x}, \lambda) \in \mathcal{X} := \mathbf{dom}\, F \times \Lambda^*.$$

- Saddle-point $(\widehat{\mathbf{x}}, \widehat{\lambda}) \in \mathcal{X}$:

$$\mathcal{L}(\widehat{\mathbf{x}}, \lambda) \le \mathcal{L}(\widehat{\mathbf{x}}, \widehat{\lambda}) \le \mathcal{L}(\boldsymbol{x}, \widehat{\lambda}) \quad \forall\, (\boldsymbol{x}, \lambda) \in \mathcal{X}$$

- Monotone inclusion

$$0 \in M(\widehat{\boldsymbol{x}}, \widehat{\lambda}), \quad M(\boldsymbol{x}, \lambda) = \begin{pmatrix} \partial F(\boldsymbol{x}) + \boldsymbol{A}^\top \lambda \\ b - \boldsymbol{A}\boldsymbol{x} \end{pmatrix}$$

Applications

Many variational/optimization problems are related to (1)/(2)/(3):

- Image processing
  - Image denoising: TV-based model, ROF
  - Image deconvolution
  - ...
- Dynamical optimal transport/Benamou–Brenier problem
- Sparse regression: Lasso, least absolute deviation (LAD)
- ...

Existing (Lagrangian-based) methods

- Augmented Lagrangian method (ALM) for (3):

$$\boldsymbol{x}_{k+1} = \operatorname*{argmin}_{\boldsymbol{x}} \mathcal{L}_\sigma(\boldsymbol{x}, \lambda_k), \quad \lambda_{k+1} = \lambda_k + \sigma(\boldsymbol{A}\boldsymbol{x}_{k+1} - b)$$

  with $\mathcal{L}_\sigma(\boldsymbol{x}, \lambda) := \mathcal{L}(\boldsymbol{x}, \lambda) + \sigma/2 \left\| \boldsymbol{A}\boldsymbol{x} - b \right\|^2, \quad \sigma > 0.$

  - Hestenes (1969) and Powell (1969)
  - Dual formulation = Proximal point algorithm (Rockafellar)
  - Uzawa method
  - Not easy to update $\boldsymbol{x}_{k+1} = (u_{k+1}, v_{k+1})$

Existing (Lagrangian-based) methods

- Augmented Lagrangian method (ALM) for (3):

$$\mathbf{x}_{k+1} = \operatorname*{argmin}_{\mathbf{x}} \mathcal{L}_\sigma(\mathbf{x}, \lambda_k), \quad \lambda_{k+1} = \lambda_k + \sigma(\boldsymbol{A}\mathbf{x}_{k+1} - b)$$

with $\mathcal{L}_\sigma(\mathbf{x}, \lambda) := \mathcal{L}(\mathbf{x}, \lambda) + \sigma/2 \left\|\boldsymbol{A}\mathbf{x} - b\right\|^2, \quad \sigma > 0.$

  - Hestenes (1969) and Powell (1969)
  - Dual formulation = Proximal point algorithm (Rockafellar)
  - Uzawa method
  - Not easy to update $\mathbf{x}_{k+1} = (u_{k+1}, v_{k+1})$

- Alternating direction method of multipliers (ADMM):

$$\left.\begin{aligned} u_{k+1} &= \operatorname*{argmin}_{u} \mathcal{L}_\sigma(u, v_k, \lambda_k) \\ v_{k+1} &= \operatorname*{argmin}_{v} \mathcal{L}_\sigma(u_{k+1}, v, \lambda_k) \end{aligned}\right\} \text{Decouple } u \text{ and } v$$

$$\lambda_{k+1} = \lambda_k + \sigma(Au_{k+1} + Bv_{k+1} - b)$$

  - Numerical solution of PDEs from mechanics, physics and differential geometry (Glowinski, 1976/2014)

- Popular application to image processing, statistical learning, data mining (Boyd et al., *Found Trends Mach Learn*, 2010)
- Successive computation, similar with Alternating Direction Method (ADM) and Gauss-Seidel iteration
- Dual formulation = DRSM
- Convergence rate:

$$O(1/\sqrt{k}), \quad O(1/k)(\textbf{Ergodic}), \quad O(1/k^2)(\mu_f + \mu_g > 0, \textbf{Ergodic})$$

- Popular application to image processing, statistical learning, data mining (Boyd et al., *Found Trends Mach Learn*, 2010)
- Successive computation, similar with Alternating Direction Method (ADM) and Gauss-Seidel iteration
- Dual formulation = DRSM
- Convergence rate:

$$O(1/\sqrt{k}), \quad O(1/k)(\textbf{Ergodic}), \quad O(1/k^2)(\mu_f + \mu_g > 0, \textbf{Ergodic})$$

- "**Ergodic**" means average of historical iterates:

$$\bar{\mathbf{x}}_k = \sum_{i=1}^{k} \alpha_i \mathbf{x}_i, \quad \sum_{i=1}^{k} \alpha_i = 1$$

- Popular application to image processing, statistical learning, data mining (Boyd et al., *Found Trends Mach Learn*, 2010)
- Successive computation, similar with Alternating Direction Method (ADM) and Gauss-Seidel iteration
- Dual formulation = DRSM
- Convergence rate:

$$O(1/\sqrt{k}), \quad O(1/k)(\textbf{Ergodic}), \quad O(1/k^2)(\mu_f + \mu_g > 0, \textbf{Ergodic})$$

- "**Ergodic**" means average of historical iterates:

$$\bar{\mathbf{x}}_k = \sum_{i=1}^{k} \alpha_i \mathbf{x}_i, \quad \sum_{i=1}^{k} \alpha_i = 1$$

- Primal-dual solvers for (2):
  - Arrow–Huricz, PDHG (Zhu and Chan, 2008)
  - Chambolle–Pock's method (*JMIV, 2011*)
  - Tight connection with ADMM
  - Ergodic convergence rate $O(1/k)$

Acceleration and nonergodic rate

- For unconstrained problem: $\min f(x)$

$$\begin{cases} O(1/k) & \mu_f = 0 \\ O(e^{-k/\mathrm{cond}(f)}) & \mu_f > 0 \end{cases} \Longrightarrow \begin{cases} O(1/k^2) & \mu_f = 0 \\ O(e^{-k/\sqrt{\mathrm{cond}(f)}}) & \mu_f > 0 \end{cases}$$

Acceleration and nonergodic rate

- For unconstrained problem: $\min f(x)$

$$\begin{cases} O(1/k) & \mu_f = 0 \\ O(e^{-k/\mathrm{cond}(f)}) & \mu_f > 0 \end{cases} \implies \begin{cases} O(1/k^2) & \mu_f = 0 \\ O(e^{-k/\sqrt{\mathrm{cond}(f)}}) & \mu_f > 0 \end{cases}$$

- **Why nonergodic?**
  Ensure sparsity or low-rankness; see Li and Lin (JSC,2019), Tran-Dinh and Zhu (SIOPT, 2020)

Acceleration and nonergodic rate

- For unconstrained problem: $\min f(x)$

$$\begin{cases} O(1/k) & \mu_f = 0 \\ O(e^{-k/\mathrm{cond}(f)}) & \mu_f > 0 \end{cases} \Longrightarrow \begin{cases} O(1/k^2) & \mu_f = 0 \\ O(e^{-k/\sqrt{\mathrm{cond}(f)}}) & \mu_f > 0 \end{cases}$$

- **Why nonergodic?**
  Ensure sparsity or low-rankness; see Li and Lin (JSC,2019), Tran-Dinh and Zhu (SIOPT, 2020)

- For problem (1), accelerated ADMM with nonergodic rate:
  - For $\mu_f = \mu_g = 0$
    - Li and Lin (JSC, 2017): $O(1/k)$
    - Ouyang et al. (SIIS, 2015): $O\left(\frac{L}{k^2} + \frac{\|A\|}{k}\right)$ (mixed-type)

Acceleration and nonergodic rate

- For unconstrained problem: $\min f(x)$

$$\begin{cases} O(1/k) & \mu_f = 0 \\ O(e^{-k/\mathrm{cond}(f)}) & \mu_f > 0 \end{cases} \Longrightarrow \begin{cases} O(1/k^2) & \mu_f = 0 \\ O(e^{-k/\sqrt{\mathrm{cond}(f)}}) & \mu_f > 0 \end{cases}$$

- **Why nonergodic?**
  Ensure sparsity or low-rankness; see Li and Lin (JSC,2019), Tran-Dinh and Zhu (SIOPT, 2020)

- For problem (1), accelerated ADMM with nonergodic rate:
  - For $\mu_f = \mu_g = 0$
    - Li and Lin (JSC, 2017): $O(1/k)$
    - Ouyang et al. (SIIS, 2015): $O\left(\frac{L}{k^2} + \frac{\|A\|}{k}\right)$ (mixed-type)
  - For $\mu_f + \mu_g > 0$: $O(1/k^2)$
    - Tran-Dinh et al.(SIOPT, 2020): semi-ergodic rate
    - Sabach and Teboulle (SIOPT, 2022)
    - Zhang et al. (arXiv:2206.05088, 2022)
    - He et al. (arXiv:2310.16404, 2023)

Main contribution

- A continuous ODE-based framework:
  - A family of accelerated ADMM from a systematic way
  - A unified Lyapunov analysis approach
  - Sharp mixed-type estimate and nonergodic rates
  - Both convex ($\mu_f = \mu_g = 0$) and (partially) strongly convex ($\mu_f + \mu_g > 0$)

Main contribution

- A continuous ODE-based framework:
  - A family of accelerated ADMM from a systematic way
  - A unified Lyapunov analysis approach
  - Sharp mixed-type estimate and nonergodic rates
  - Both convex ($\mu_f = \mu_g = 0$) and (partially) strongly convex ($\mu_f + \mu_g > 0$)
- Applications to sparse regression: fast nonergodic convergence and sparsity maintaining

PDHG and CP

- PDHG/Arrow–Hurwicz algorithm [3]

$$\begin{aligned} \boldsymbol{x}_{k+1} &= \operatorname*{argmin}_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}, \lambda_k) + \frac{1}{2\tau} \|\boldsymbol{x} - \boldsymbol{x}_k\|^2 \\ \lambda_{k+1} &= \operatorname*{argmax}_{\lambda} \mathcal{L}(\boldsymbol{x}_{k+1}, \lambda) - \frac{1}{2\tau} \|\lambda - \lambda_k\|^2 \end{aligned}$$

[3] M. Zhu and T. F. Chan. An efficient primal-dual hybrid gradient algorithm for total variation image restoration. CAM Report 08-34, 2008.

PDHG and CP

- PDHG/Arrow–Hurwicz algorithm [3]

$$
\begin{aligned}
\boldsymbol{x}_{k+1} &= \operatorname*{argmin}_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}, \lambda_k) + \frac{1}{2\tau} \left\|\boldsymbol{x} - \boldsymbol{x}_k\right\|^2 \\
\lambda_{k+1} &= \operatorname*{argmax}_{\lambda} \mathcal{L}(\boldsymbol{x}_{k+1}, \lambda) - \frac{1}{2\tau} \left\|\lambda - \lambda_k\right\|^2
\end{aligned}
$$

  - $0 \in S\,[\boldsymbol{z}_{k+1} - \boldsymbol{z}_k)] + \tau M(\boldsymbol{z}_{k+1}),\ S = \begin{pmatrix} I & -\tau \boldsymbol{A}^\top \\ O & I \end{pmatrix}$
  - Preconditioned PPA without symmetric property [4]

[3] M. Zhu and T. F. Chan. An efficient primal-dual hybrid gradient algorithm for total variation image restoration. CAM Report 08-34, 2008.

[4] B. He and X. Yuan. Convergence analysis of primal-dual algorithms for a saddle-point problem: from contraction perspective. SIIS, 5(1):119–149, 2012.

PDHG and CP

- PDHG/Arrow–Hurwicz algorithm [3]

$$\boldsymbol{x}_{k+1} = \underset{\boldsymbol{x}}{\operatorname{argmin}}\, \mathcal{L}(\boldsymbol{x}, \lambda_k) + \frac{1}{2\tau} \|\boldsymbol{x} - \boldsymbol{x}_k\|^2$$

$$\lambda_{k+1} = \underset{\lambda}{\operatorname{argmax}}\, \mathcal{L}(\boldsymbol{x}_{k+1}, \lambda) - \frac{1}{2\tau} \|\lambda - \lambda_k\|^2$$

  - $0 \in S\left[\boldsymbol{z}_{k+1} - \boldsymbol{z}_k)\right] + \tau M(\boldsymbol{z}_{k+1}),\ S = \begin{pmatrix} I & -\tau \boldsymbol{A}^\top \\ O & I \end{pmatrix}$
  - Preconditioned PPA without symmetric property [4]
  - Asymmetric $\Rightarrow$ Diverge! Counter examples in He et al. *JIMV* (2022)

---

[3] M. Zhu and T. F. Chan. An efficient primal-dual hybrid gradient algorithm for total variation image restoration. CAM Report 08-34, 2008.

[4] B. He and X. Yuan. Convergence analysis of primal-dual algorithms for a saddle-point problem: from contraction perspective. SIIS, 5(1):119–149, 2012.

PDHG and CP

- PDHG/Arrow–Hurwicz algorithm [3]

$$\begin{aligned} \boldsymbol{x}_{k+1} &= \operatorname*{argmin}_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}, \lambda_k) + \frac{1}{2\tau} \|\boldsymbol{x} - \boldsymbol{x}_k\|^2 \\ \lambda_{k+1} &= \operatorname*{argmax}_{\lambda} \mathcal{L}(\boldsymbol{x}_{k+1}, \lambda) - \frac{1}{2\tau} \|\lambda - \lambda_k\|^2 \end{aligned}$$

  - $0 \in S\left[\boldsymbol{z}_{k+1} - \boldsymbol{z}_k)\right] + \tau M(\boldsymbol{z}_{k+1}),\ S = \begin{pmatrix} I & -\tau \boldsymbol{A}^\top \\ O & I \end{pmatrix}$
  - Preconditioned PPA without symmetric property [4]
  - Asymmetric $\Rightarrow$ Diverge! Counter examples in He et al. *JIMV* (2022)
- Chambolle-Pock (CP) method [5]

$$\begin{aligned} \boldsymbol{x}_{k+1} &= \operatorname*{argmin}_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}, \lambda_k) + \frac{1}{2\tau} \|\boldsymbol{x} - \boldsymbol{x}_k\|^2 \\ \lambda_{k+1} &= \operatorname*{argmax}_{\lambda} \mathcal{L}(\boldsymbol{x}_{k+1} + \boldsymbol{x}_{k+1} - \boldsymbol{x}_k, \lambda) - \frac{1}{2\tau} \|\lambda - \lambda_k\|^2 \end{aligned}$$

[3] M. Zhu and T. F. Chan. An efficient primal-dual hybrid gradient algorithm for total variation image restoration. CAM Report 08-34, 2008.

[4] B. He and X. Yuan. Convergence analysis of primal-dual algorithms for a saddle-point problem: from contraction perspective. SIIS, 5(1):119–149, 2012.

PDHG and CP

- PDHG/Arrow–Hurwicz algorithm [3]

$$x_{k+1} = \operatorname*{argmin}_{x} \mathcal{L}(x, \lambda_k) + \frac{1}{2\tau}\|x - x_k\|^2$$

$$\lambda_{k+1} = \operatorname*{argmax}_{\lambda} \mathcal{L}(x_{k+1}, \lambda) - \frac{1}{2\tau}\|\lambda - \lambda_k\|^2$$

  - $0 \in S\,[z_{k+1} - z_k)] + \tau M(z_{k+1}),\ S = \begin{pmatrix} I & -\tau A^\top \\ O & I \end{pmatrix}$
  - Preconditioned PPA without symmetric property [4]
  - Asymmetric $\Rightarrow$ Diverge! Counter examples in He et al. *JIMV* (2022)

- Chambolle-Pock (CP) method [5]

$$x_{k+1} = \operatorname*{argmin}_{x} \mathcal{L}(x, \lambda_k) + \frac{1}{2\tau}\|x - x_k\|^2$$

$$\lambda_{k+1} = \operatorname*{argmax}_{\lambda} \mathcal{L}(x_{k+1} + x_{k+1} - x_k, \lambda) - \frac{1}{2\tau}\|\lambda - \lambda_k\|^2$$

  - $0 \in S\,[z_{k+1} - z_k)] + \tau M(z_{k+1}),\ S = \begin{pmatrix} I & -\tau A^\top \\ -\tau A & I \end{pmatrix}$

[3] M. Zhu and T. F. Chan. An efficient primal-dual hybrid gradient algorithm for total variation image restoration. CAM Report 08-34, 2008.

[4] B. He and X. Yuan. Convergence analysis of primal-dual algorithms for a saddle-point problem: from contraction perspective. SIIS, 5(1):119–149, 2012.

Primal-dual flow

- Rewriting CP method as a semi-implicit discretization

$$
\begin{aligned}
\frac{\boldsymbol{x}_{k+1} - \boldsymbol{x}_k}{\tau} &\in -\partial_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}_{k+1}, \lambda_k) \\
\frac{\lambda_{k+1} - \lambda_k}{\tau} &= \nabla_\lambda \mathcal{L}(\boldsymbol{x}_{k+1} + \boldsymbol{x}_{k+1} - \boldsymbol{x}_k, \lambda_{k+1})
\end{aligned}
$$

Primal-dual flow

- Rewriting CP method as a semi-implicit discretization

$$
\begin{aligned}
\frac{\boldsymbol{x}_{k+1}-\boldsymbol{x}_k}{\tau} &\in -\partial_{\boldsymbol{x}}\mathcal{L}(\boldsymbol{x}_{k+1},\lambda_k)\\
\frac{\lambda_{k+1}-\lambda_k}{\tau} &= \nabla_\lambda\mathcal{L}(\boldsymbol{x}_{k+1}+\boldsymbol{x}_{k+1}-\boldsymbol{x}_k,\lambda_{k+1})
\end{aligned}
$$

- We proposed a novel primal-dual flow [6]

$$
\boxed{\begin{aligned}
&\gamma\boldsymbol{x}' \in -\partial_{\boldsymbol{x}}\mathcal{L}(\boldsymbol{x},\lambda)\\
&\theta\lambda' = \nabla_\lambda\mathcal{L}(\boldsymbol{x}+\boldsymbol{x}',\lambda)
\end{aligned}} \qquad \textbf{(PD Flow)}
$$

[6] L. A primal-dual flow for affine constrained convex optimization. *ESAIM:COCV* (2022)

Primal-dual flow

- Rewriting CP method as a semi-implicit discretization

$$
\begin{aligned}
\frac{\boldsymbol{x}_{k+1}-\boldsymbol{x}_k}{\tau} &\in -\partial_{\boldsymbol{x}}\mathcal{L}(\boldsymbol{x}_{k+1},\lambda_k)\\
\frac{\lambda_{k+1}-\lambda_k}{\tau} &= \nabla_\lambda\mathcal{L}(\boldsymbol{x}_{k+1}+\boldsymbol{x}_{k+1}-\boldsymbol{x}_k,\lambda_{k+1})
\end{aligned}
$$

- We proposed a novel primal-dual flow [6]

$$
\boxed{\begin{aligned}
&\gamma\boldsymbol{x}' \in -\partial_{\boldsymbol{x}}\mathcal{L}(\boldsymbol{x},\lambda)\\
&\theta\lambda' = \nabla_\lambda\mathcal{L}(\boldsymbol{x}+\boldsymbol{x}',\lambda)
\end{aligned}} \qquad \textbf{(PD Flow)}
$$

- Scaling parameters $\theta' = -\theta$ and $\gamma' = \mu - \gamma$ with $\mu = \mu_F$

[6] L. A primal-dual flow for affine constrained convex optimization. *ESAIM:COCV* (2022)

Primal-dual flow

- Rewriting CP method as a semi-implicit discretization

$$\begin{aligned}\frac{\boldsymbol{x}_{k+1}-\boldsymbol{x}_k}{\tau} &\in -\partial_{\boldsymbol{x}}\mathcal{L}(\boldsymbol{x}_{k+1},\lambda_k)\\ \frac{\lambda_{k+1}-\lambda_k}{\tau} &= \nabla_\lambda\mathcal{L}(\boldsymbol{x}_{k+1}+\boldsymbol{x}_{k+1}-\boldsymbol{x}_k,\lambda_{k+1})\end{aligned}$$

- We proposed a novel primal-dual flow [6]

$$\boxed{\begin{aligned}\gamma\boldsymbol{x}' &\in -\partial_{\boldsymbol{x}}\mathcal{L}(\boldsymbol{x},\lambda)\\ \theta\lambda' &= \nabla_\lambda\mathcal{L}(\boldsymbol{x}+\boldsymbol{x}',\lambda)\end{aligned}} \qquad \textbf{(PD Flow)}$$

- Scaling parameters $\theta' = -\theta$ and $\gamma' = \mu - \gamma$ with $\mu = \mu_F$
- Exponential decay of the Lyapunov function

$$\mathcal{E}(\boldsymbol{x},\lambda) := \mathcal{L}(\boldsymbol{x},\widehat{\lambda}) - \mathcal{L}(\widehat{\boldsymbol{x}},\lambda) + \frac{\gamma}{2}\left\|\boldsymbol{x}-\widehat{\boldsymbol{x}}\right\|^2 + \frac{\theta}{2}\|\lambda-\widehat{\lambda}\|^2$$

[6] L. A primal-dual flow for affine constrained convex optimization. *ESAIM:COCV* (2022)

Accelerated primal-dual flow

- **Main idea**: Combining primal-dual form with acceleration

[7] L. and L. Chen. From differential equation solvers to accelerated first-order methods for convex optimization. MAPR, 195: 735–781, 2022.

Accelerated primal-dual flow

- **Main idea**: Combining primal-dual form with acceleration
- Nesterov accelerated gradient flow for $\min F(\boldsymbol{x})$ [7]

$$\gamma \boldsymbol{x}'' + (\mu + \gamma)\boldsymbol{x}' + \partial F(\boldsymbol{x}) \ni 0 \qquad \text{(NAG flow)}$$

[7] L. and L. Chen. From differential equation solvers to accelerated first-order methods for convex optimization. MAPR, 195: 735–781, 2022.

Accelerated primal-dual flow

- **Main idea**: Combining primal-dual form with acceleration
- Nesterov accelerated gradient flow for $\min F(\boldsymbol{x})$ [7]

$$\gamma \boldsymbol{x}'' + (\mu + \gamma)\boldsymbol{x}' + \partial F(\boldsymbol{x}) \ni 0 \qquad \text{(NAG flow)}$$

- Combining NAG-flow with **(PD Flow)** gives

$$\begin{aligned} \gamma \mathbf{x}' &\in -\,\partial_{\mathbf{x}} \mathcal{L}(\boldsymbol{x}, \lambda) \quad \text{[Primal descent]} \\ \theta \lambda' &= \nabla_\lambda \mathcal{L}(\boldsymbol{x} + \mathbf{x}', \lambda) \quad \text{[Dual ascent]} \end{aligned}$$

[7] L. and L. Chen. From differential equation solvers to accelerated first-order methods for convex optimization. MAPR, 195: 735–781, 2022.

Accelerated primal-dual flow

- **Main idea**: Combining primal-dual form with acceleration
- Nesterov accelerated gradient flow for $\min F(\boldsymbol{x})$ [7]

$$\gamma \boldsymbol{x}'' + (\mu + \gamma)\boldsymbol{x}' + \partial F(\boldsymbol{x}) \ni 0 \qquad \text{(NAG flow)}$$

- Combining NAG-flow with **(PD Flow)** gives

$$\gamma \mathbf{x}'' + (\mu + \gamma)\mathbf{x}' \in -\,\partial_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}, \lambda) \quad \text{[Accelerated primal descent]}$$
$$\theta \lambda' = \nabla_\lambda \mathcal{L}(\mathbf{x} + \mathbf{x}', \lambda) \quad \text{[Dual ascent]}$$

[7] L. and L. Chen. From differential equation solvers to accelerated first-order methods for convex optimization. MAPR, 195: 735–781, 2022.

Accelerated primal-dual flow

- **Main idea**: Combining primal-dual form with acceleration
- Nesterov accelerated gradient flow for $\min F(\boldsymbol{x})$ [7]

$$\gamma \boldsymbol{x}'' + (\mu + \gamma)\boldsymbol{x}' + \partial F(\boldsymbol{x}) \ni 0 \qquad \text{(NAG flow)}$$

- Combining NAG-flow with **(PD Flow)** gives

$$\begin{aligned} \gamma \mathbf{x}'' + (\mu + \gamma)\mathbf{x}' &\in -\partial_{\mathbf{x}} \mathcal{L}(\boldsymbol{x}, \lambda) \quad \text{[Accelerated primal descent]} \\ \theta \lambda' &= \nabla_\lambda \mathcal{L}(\mathbf{x} + \mathbf{x}', \lambda) \quad \text{[Dual ascent]} \end{aligned}$$

- Introduce $\bar{\mathbf{x}} = \mathbf{x} + \mathbf{x}'$ to obtain

$$\boxed{\begin{aligned} \mathbf{x}' &= \bar{\mathbf{x}} - \mathbf{x} \\ \gamma \bar{\mathbf{x}}' &\in -\partial_{\mathbf{x}} \mathcal{L}(\boldsymbol{x}, \lambda) + \mu(\boldsymbol{x} - \bar{\mathbf{x}}) \\ \theta \lambda' &= \nabla_\lambda \mathcal{L}(\bar{\mathbf{x}}, \lambda) \end{aligned}} \qquad \textbf{(APD Flow)}$$

  - First-order system is convenient for discretization and analysis than second-order ODE

[7] L. and L. Chen. From differential equation solvers to accelerated first-order methods for convex optimization. MAPR, 195: 735–781, 2022.

Two block APD flow

- Recall that $\mu = \mu_F = \min\{\mu_f, \mu_g\}$, and $\mu = 0$ when $\mu_f \mu_g = 0$

Two block APD flow

- Recall that $\mu = \mu_F = \min\{\mu_f,\, \mu_g\}$, and $\mu = 0$ when $\mu_f \mu_g = 0$
- For two block problem (1), we introduce **block diagonal scaling parameters**:

$$\gamma \implies \Gamma = \operatorname{diag}\left(\gamma I, \beta I\right), \quad \mu \implies \boldsymbol{\mu} = \operatorname{diag}\left(\mu_f I, \mu_g I\right)$$

$$\gamma' = \mu - \gamma \implies \Gamma' = \boldsymbol{\mu} - \Gamma$$

Two block APD flow

- Recall that $\mu = \mu_F = \min\{\mu_f,\, \mu_g\}$, and $\mu = 0$ when $\mu_f \mu_g = 0$
- For two block problem (1), we introduce **block diagonal scaling parameters**:

$$\gamma \implies \Gamma = \operatorname{diag}(\gamma I, \beta I), \quad \mu \implies \boldsymbol{\mu} = \operatorname{diag}(\mu_f I, \mu_g I)$$

$$\gamma' = \mu - \gamma \implies \Gamma' = \boldsymbol{\mu} - \Gamma$$

- Two block APD flow

$$\boxed{\begin{aligned} \boldsymbol{x}' &= \bar{\boldsymbol{x}} - \boldsymbol{x} \\ \Gamma \bar{\boldsymbol{x}}' &\in -\partial_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}, \lambda) + \boldsymbol{\mu}(\boldsymbol{x} - \bar{\boldsymbol{x}}) \\ \theta \lambda' &= \nabla_\lambda \mathcal{L}(\bar{\boldsymbol{x}}, \lambda) \end{aligned}} \qquad \textbf{(2block-APD Flow)}$$

Two block APD flow

- Recall that $\mu = \mu_F = \min\{\mu_f, \mu_g\}$, and $\mu = 0$ when $\mu_f \mu_g = 0$
- For two block problem (1), we introduce **block diagonal scaling parameters**:

$$\gamma \implies \Gamma = \operatorname{diag}(\gamma I, \beta I), \quad \mu \implies \boldsymbol{\mu} = \operatorname{diag}(\mu_f I, \mu_g I)$$

$$\gamma' = \mu - \gamma \implies \Gamma' = \boldsymbol{\mu} - \Gamma$$

- Two block APD flow

$$\boxed{\begin{aligned} &\boldsymbol{x}' = \bar{\mathbf{x}} - \boldsymbol{x} \\ &\Gamma \bar{\mathbf{x}}' \in -\partial_{\mathbf{x}} \mathcal{L}(\mathbf{x}, \lambda) + \boldsymbol{\mu}(\boldsymbol{x} - \bar{\mathbf{x}}) \\ &\theta \lambda' = \nabla_\lambda \mathcal{L}(\bar{\mathbf{x}}, \lambda) \end{aligned}} \qquad \textbf{(2block-APD Flow)}$$

- Exponential decay of

$$\mathcal{E}(\mathbf{x}, \lambda) := \mathcal{L}(\mathbf{x}, \widehat{\lambda}) - \mathcal{L}(\widehat{\mathbf{x}}, \lambda) + \frac{1}{2}\|\bar{\mathbf{x}} - \widehat{\mathbf{x}}\|_\Gamma^2 + \frac{\theta}{2}\|\lambda - \widehat{\lambda}\|^2$$

A discretization template

- Denote the step size $\alpha_k > 0$ and consider

$$
\begin{aligned}
\frac{\mathbf{x}_{k+1} - \mathbf{x}_k}{\alpha_k} &= \bar{\mathbf{x}}_{k+1} - \mathbf{x}_{k+1} \\
\Gamma_k \frac{\bar{\mathbf{x}}_{k+1} - \bar{\mathbf{x}}_k}{\alpha_k} &\in -\partial_{\mathbf{x}} \mathcal{L}(\mathbf{x}_{k+1}, \bar{\lambda}_{k+1}) + \boldsymbol{\mu}(\mathbf{x}_{k+1} - \bar{\mathbf{x}}_{k+1}) \\
\theta_k \frac{\lambda_{k+1} - \lambda_k}{\alpha_k} &= \boldsymbol{A} \bar{\mathbf{x}}_{k+1} - b
\end{aligned}
$$

A discretization template

- Denote the step size $\alpha_k > 0$ and consider

$$
\begin{aligned}
\frac{\mathbf{x}_{k+1} - \mathbf{x}_k}{\alpha_k} &= \bar{\mathbf{x}}_{k+1} - \mathbf{x}_{k+1} \\
\Gamma_k \frac{\bar{\mathbf{x}}_{k+1} - \bar{\mathbf{x}}_k}{\alpha_k} &\in -\partial_{\mathbf{x}} \mathcal{L}(\mathbf{x}_{k+1}, {\color{red}\bar{\lambda}_{k+1}}) + \boldsymbol{\mu}(\mathbf{x}_{k+1} - \bar{\mathbf{x}}_{k+1}) \\
\theta_k \frac{\lambda_{k+1} - \lambda_k}{\alpha_k} &= \boldsymbol{A}\bar{\mathbf{x}}_{k+1} - b
\end{aligned}
$$

- Parameter equation

$$
\frac{\Gamma_{k+1} - \Gamma_k}{\alpha_k} = \boldsymbol{\mu} - \Gamma_{k+1}, \quad \frac{\theta_{k+1} - \theta_k}{\alpha_k} = -\theta_{k+1}
$$

A discretization template

- Denote the step size $\alpha_k > 0$ and consider

$$
\begin{aligned}
\frac{\boldsymbol{x}_{k+1} - \boldsymbol{x}_k}{\alpha_k} &= \bar{\boldsymbol{x}}_{k+1} - \boldsymbol{x}_{k+1} \\
\Gamma_k \frac{\bar{\boldsymbol{x}}_{k+1} - \bar{\boldsymbol{x}}_k}{\alpha_k} &\in - \partial_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}_{k+1}, \bar{\lambda}_{k+1}) + \boldsymbol{\mu}(\boldsymbol{x}_{k+1} - \bar{\boldsymbol{x}}_{k+1}) \\
\theta_k \frac{\lambda_{k+1} - \lambda_k}{\alpha_k} &= \boldsymbol{A}\bar{\boldsymbol{x}}_{k+1} - b
\end{aligned}
$$

- Parameter equation

$$
\frac{\Gamma_{k+1} - \Gamma_k}{\alpha_k} = \boldsymbol{\mu} - \Gamma_{k+1}, \quad \frac{\theta_{k+1} - \theta_k}{\alpha_k} = -\theta_{k+1}
$$

- $\bar{\lambda}_{k+1}$ is crucial to decouple $\boldsymbol{x}_{k+1}$ and $\bar{\boldsymbol{x}}_{k+1}$:

A discretization template

- Denote the step size $\alpha_k > 0$ and consider

$$
\begin{aligned}
\frac{\boldsymbol{x}_{k+1} - \boldsymbol{x}_k}{\alpha_k} &= \bar{\boldsymbol{x}}_{k+1} - \boldsymbol{x}_{k+1} \\
\Gamma_k \frac{\bar{\boldsymbol{x}}_{k+1} - \bar{\boldsymbol{x}}_k}{\alpha_k} &\in -\partial_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}_{k+1}, \bar{\lambda}_{k+1}) + \boldsymbol{\mu}(\boldsymbol{x}_{k+1} - \bar{\boldsymbol{x}}_{k+1}) \\
\theta_k \frac{\lambda_{k+1} - \lambda_k}{\alpha_k} &= \boldsymbol{A}\bar{\boldsymbol{x}}_{k+1} - b
\end{aligned}
$$

- Parameter equation

$$
\frac{\Gamma_{k+1} - \Gamma_k}{\alpha_k} = \boldsymbol{\mu} - \Gamma_{k+1}, \quad \frac{\theta_{k+1} - \theta_k}{\alpha_k} = -\theta_{k+1}
$$

- $\bar{\lambda}_{k+1}$ is crucial to decouple $\boldsymbol{x}_{k+1}$ and $\bar{\boldsymbol{x}}_{k+1}$:
  - $\bar{\lambda}_{k+1} = \lambda_{k+1}$: fully coupled $\Longrightarrow$ nonlinear $\&$ nonsmooth ✘

A discretization template

- Denote the step size $\alpha_k > 0$ and consider

$$
\begin{aligned}
\frac{\boldsymbol{x}_{k+1}-\boldsymbol{x}_k}{\alpha_k} &= \bar{\boldsymbol{x}}_{k+1}-\boldsymbol{x}_{k+1}\\
\Gamma_k\frac{\bar{\boldsymbol{x}}_{k+1}-\bar{\boldsymbol{x}}_k}{\alpha_k} &\in -\partial_{\boldsymbol{x}}\mathcal{L}(\boldsymbol{x}_{k+1},\bar{\lambda}_{k+1})+\boldsymbol{\mu}(\boldsymbol{x}_{k+1}-\bar{\boldsymbol{x}}_{k+1})\\
\theta_k\frac{\lambda_{k+1}-\lambda_k}{\alpha_k} &= \boldsymbol{A}\bar{\boldsymbol{x}}_{k+1}-b
\end{aligned}
$$

- Parameter equation

$$
\frac{\Gamma_{k+1}-\Gamma_k}{\alpha_k} = \boldsymbol{\mu}-\Gamma_{k+1},\quad \frac{\theta_{k+1}-\theta_k}{\alpha_k} = -\theta_{k+1}
$$

- $\bar{\lambda}_{k+1}$ is crucial to decouple $\boldsymbol{x}_{k+1}$ and $\bar{\boldsymbol{x}}_{k+1}$:
  - $\bar{\lambda}_{k+1}=\lambda_{k+1}$: fully coupled $\Longrightarrow$ nonlinear & nonsmooth ✘
  - $\bar{\lambda}_{k+1}\sim(u_{k+1},\bar{u}_{k+1})$: weakly coupled $\Longrightarrow$ splitting ✔

A discretization template

- Denote the step size $\alpha_k > 0$ and consider

$$
\begin{aligned}
\frac{\boldsymbol{x}_{k+1} - \boldsymbol{x}_k}{\alpha_k} &= \bar{\boldsymbol{x}}_{k+1} - \boldsymbol{x}_{k+1} \\
\Gamma_k \frac{\bar{\boldsymbol{x}}_{k+1} - \bar{\boldsymbol{x}}_k}{\alpha_k} &\in -\partial_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}_{k+1}, \bar{\lambda}_{k+1}) + \boldsymbol{\mu}(\boldsymbol{x}_{k+1} - \bar{\boldsymbol{x}}_{k+1}) \\
\theta_k \frac{\lambda_{k+1} - \lambda_k}{\alpha_k} &= \boldsymbol{A}\bar{\boldsymbol{x}}_{k+1} - b
\end{aligned}
$$

- Parameter equation

$$
\frac{\Gamma_{k+1} - \Gamma_k}{\alpha_k} = \boldsymbol{\mu} - \Gamma_{k+1}, \quad \frac{\theta_{k+1} - \theta_k}{\alpha_k} = -\theta_{k+1}
$$

- $\bar{\lambda}_{k+1}$ is crucial to decouple $\boldsymbol{x}_{k+1}$ and $\bar{\boldsymbol{x}}_{k+1}$:
  - $\bar{\lambda}_{k+1} = \lambda_{k+1}$: fully coupled $\Longrightarrow$ nonlinear & nonsmooth ✘
  - $\bar{\lambda}_{k+1} \sim (u_{k+1}, \bar{u}_{k+1})$: weakly coupled $\Longrightarrow$ splitting ✔
  - $\bar{\lambda}_{k+1} \sim (v_{k+1}, \bar{v}_{k+1})$: weakly coupled $\Longrightarrow$ splitting ✔

A discretization template

- Denote the step size $\alpha_k > 0$ and consider

$$
\begin{aligned}
\frac{\boldsymbol{x}_{k+1} - \boldsymbol{x}_k}{\alpha_k} &= \bar{\boldsymbol{x}}_{k+1} - \boldsymbol{x}_{k+1} \\
\Gamma_k \frac{\bar{\boldsymbol{x}}_{k+1} - \bar{\boldsymbol{x}}_k}{\alpha_k} &\in -\partial_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}_{k+1}, \bar{\lambda}_{k+1}) + \boldsymbol{\mu}(\boldsymbol{x}_{k+1} - \bar{\boldsymbol{x}}_{k+1}) \\
\theta_k \frac{\lambda_{k+1} - \lambda_k}{\alpha_k} &= \boldsymbol{A}\bar{\boldsymbol{x}}_{k+1} - b
\end{aligned}
$$

- Parameter equation

$$
\frac{\Gamma_{k+1} - \Gamma_k}{\alpha_k} = \boldsymbol{\mu} - \Gamma_{k+1}, \quad \frac{\theta_{k+1} - \theta_k}{\alpha_k} = -\theta_{k+1}
$$

- $\bar{\lambda}_{k+1}$ is crucial to decouple $\boldsymbol{x}_{k+1}$ and $\bar{\boldsymbol{x}}_{k+1}$:
  - $\bar{\lambda}_{k+1} = \lambda_{k+1}$: fully coupled $\Longrightarrow$ nonlinear & nonsmooth ✘
  - $\bar{\lambda}_{k+1} \sim (u_{k+1}, \bar{u}_{k+1})$: weakly coupled $\Longrightarrow$ splitting ✔
  - $\bar{\lambda}_{k+1} \sim (v_{k+1}, \bar{v}_{k+1})$: weakly coupled $\Longrightarrow$ splitting ✔
  - $\bar{\lambda}_{k+1} \sim (u_k, v_k)$: decoupled $\Longrightarrow$ splitting + parallelization ✔

One iteration analysis

Discrete Lyapunov function (From the continuous one)

$$\mathcal{E}_k := \mathcal{L}(\boldsymbol{x}_k, \widehat{\lambda}) - \mathcal{L}(\widehat{\boldsymbol{x}}, \lambda_k) + \frac{1}{2} \left\| \bar{\boldsymbol{x}}_k - \widehat{\boldsymbol{x}} \right\|_{\Gamma_k}^2 + \frac{\theta_k}{2} \left\| \lambda_k - \widehat{\lambda} \right\|^2$$

Lemma 1

$$\mathcal{E}_{k+1} - \mathcal{E}_k \leq -\alpha_k \mathcal{E}_{k+1} + \frac{\theta_k}{2} \left\| \lambda_{k+1} - \bar{\lambda}_{k+1} \right\|^2 - \frac{1}{2} \left\| \bar{\boldsymbol{x}}_{k+1} - \bar{\boldsymbol{x}}_k \right\|_{\Gamma_k}^2$$

One iteration analysis

Discrete Lyapunov function (From the continuous one)

$$\mathcal{E}_k := \mathcal{L}(\boldsymbol{x}_k, \widehat{\lambda}) - \mathcal{L}(\widehat{\mathbf{x}}, \lambda_k) + \frac{1}{2}\left\|\bar{\mathbf{x}}_k - \widehat{\mathbf{x}}\right\|_{\Gamma_k}^2 + \frac{\theta_k}{2}\left\|\lambda_k - \widehat{\lambda}\right\|^2$$

Lemma 1

$$\mathcal{E}_{k+1} - \mathcal{E}_k \leq -\alpha_k \mathcal{E}_{k+1} + \frac{\theta_k}{2}\left\|\lambda_{k+1} - \bar{\lambda}_{k+1}\right\|^2 - \frac{1}{2}\left\|\bar{\mathbf{x}}_{k+1} - \bar{\mathbf{x}}_k\right\|_{\Gamma_k}^2$$

- $\bar{\lambda}_{k+1}$ is also crucial for convergence:

One iteration analysis

Discrete Lyapunov function (From the continuous one)

$$\mathcal{E}_k := \mathcal{L}(\boldsymbol{x}_k, \widehat{\lambda}) - \mathcal{L}(\widehat{\boldsymbol{x}}, \lambda_k) + \frac{1}{2}\left\|\bar{\boldsymbol{x}}_k - \widehat{\boldsymbol{x}}\right\|_{\Gamma_k}^2 + \frac{\theta_k}{2}\left\|\lambda_k - \widehat{\lambda}\right\|^2$$

Lemma 1

$$\mathcal{E}_{k+1} - \mathcal{E}_k \leq -\alpha_k \mathcal{E}_{k+1} + \frac{\theta_k}{2}\left\|\lambda_{k+1} - \bar{\lambda}_{k+1}\right\|^2 - \frac{1}{2}\left\|\bar{\boldsymbol{x}}_{k+1} - \bar{\boldsymbol{x}}_k\right\|_{\Gamma_k}^2$$

- $\bar{\lambda}_{k+1}$ is also crucial for convergence:
  - $\bar{\lambda}_{k+1} - \lambda_{k+1} = 0$: perfect ✔

One iteration analysis

Discrete Lyapunov function (From the continuous one)

$$\mathcal{E}_k := \mathcal{L}(\boldsymbol{x}_k, \widehat{\lambda}) - \mathcal{L}(\widehat{\mathbf{x}}, \lambda_k) + \frac{1}{2}\left\|\bar{\mathbf{x}}_k - \widehat{\mathbf{x}}\right\|_{\Gamma_k}^2 + \frac{\theta_k}{2}\left\|\lambda_k - \widehat{\lambda}\right\|^2$$

Lemma 1

$$\mathcal{E}_{k+1} - \mathcal{E}_k \leq -\alpha_k \mathcal{E}_{k+1} + \frac{\theta_k}{2}\left\|\lambda_{k+1} - \bar{\lambda}_{k+1}\right\|^2 - \frac{1}{2}\left\|\bar{\boldsymbol{x}}_{k+1} - \bar{\boldsymbol{x}}_k\right\|_{\Gamma_k}^2$$

- $\bar{\lambda}_{k+1}$ is also crucial for convergence:
  - $\bar{\lambda}_{k+1} - \lambda_{k+1} = 0$: perfect ✔
  - $\bar{\lambda}_{k+1} - \lambda_{k+1} \sim (\bar{u}_{k+1} - \bar{u}_k)$: controllable ✓

One iteration analysis

Discrete Lyapunov function (From the continuous one)

$$\mathcal{E}_k := \mathcal{L}(\boldsymbol{x}_k, \widehat{\lambda}) - \mathcal{L}(\widehat{\mathbf{x}}, \lambda_k) + \frac{1}{2}\left\|\bar{\mathbf{x}}_k - \widehat{\mathbf{x}}\right\|_{\Gamma_k}^2 + \frac{\theta_k}{2}\left\|\lambda_k - \widehat{\lambda}\right\|^2$$

Lemma 1

$$\mathcal{E}_{k+1} - \mathcal{E}_k \leq -\alpha_k \mathcal{E}_{k+1} + \frac{\theta_k}{2}\left\|\lambda_{k+1} - \bar{\lambda}_{k+1}\right\|^2 - \frac{1}{2}\left\|\bar{\mathbf{x}}_{k+1} - \bar{\mathbf{x}}_k\right\|_{\Gamma_k}^2$$

- $\bar{\lambda}_{k+1}$ is also crucial for convergence:
  - $\bar{\lambda}_{k+1} - \lambda_{k+1} = 0$: perfect ✔
  - $\bar{\lambda}_{k+1} - \lambda_{k+1} \sim (\bar{u}_{k+1} - \bar{u}_k)$: controllable ✓
  - $\bar{\lambda}_{k+1} - \lambda_{k+1} \sim (\bar{v}_{k+1} - \bar{v}_k)$: controllable ✓

The first IMEX scheme

- The implicit-explicit choice $\bar{\lambda}_{k+1} = \lambda_k + \alpha_k/\theta_k(A\bar{u}_{k+1} + B\bar{v}_k - b)$

The first IMEX scheme

- The implicit-explicit choice $\bar{\lambda}_{k+1} = \lambda_k + \alpha_k/\theta_k(A\bar{u}_{k+1} + B\bar{v}_k - b)$
- Controllable gap

$$\lambda_{k+1} - \bar{\lambda}_{k+1} = \alpha_k/\theta_k A(\bar{u}_{k+1} - \bar{u}_k)$$

The first IMEX scheme

- The implicit-explicit choice $\bar{\lambda}_{k+1} = \lambda_k + \alpha_k/\theta_k(A\bar{u}_{k+1} + B\bar{v}_k - b)$
- Controllable gap

$$\lambda_{k+1} - \bar{\lambda}_{k+1} = \alpha_k/\theta_k A(\bar{u}_{k+1} - \bar{u}_k)$$

- Optimization algorithm (informal)

$$\begin{cases} u_{k+1} = \underset{u}{\operatorname{argmin}} \left\{ \mathcal{L}_{\sigma_k}(u, v_k, \widehat{\lambda}_k) + \dfrac{\eta_{f,k}}{2\alpha_k^2} \left\| u - \widetilde{u}_k \right\|^2 \right\} \\ \bar{u}_{k+1} = u_{k+1} + (u_{k+1} - u_k)/\alpha_k \\ v_{k+1} = \mathbf{prox}_{\tau_k g}(\widetilde{v}_k - \tau_k B^* \bar{\lambda}_k) \\ \bar{v}_{k+1} = v_{k+1} + (v_{k+1} - v_k)/\alpha_k \\ \lambda_{k+1} = \lambda_k + \alpha_k/\theta_k(A\bar{u}_{k+1} + B\bar{v}_{k+1} - b) \end{cases} \tag{Fast-ADMM-1}$$

with last iterate u_k, v_k, λ_k, intermediate sequence $\widehat{\lambda}_k, \bar{\lambda}_k, \widetilde{u}_k, \widetilde{v}_k$ and parameter sequence $\tau_k, \eta_{f,k}, \eta_{g,k}$.

The first IMEX scheme

- The implicit-explicit choice $\bar{\lambda}_{k+1} = \lambda_k + \alpha_k/\theta_k(A\bar{u}_{k+1} + B\bar{v}_k - b)$
- Controllable gap

$$\lambda_{k+1} - \bar{\lambda}_{k+1} = \alpha_k/\theta_k A(\bar{u}_{k+1} - \bar{u}_k)$$

- Optimization algorithm (informal)

$$\begin{cases} u_{k+1} = \underset{u}{\operatorname{argmin}} \left\{ \mathcal{L}_{\sigma_k}(u, v_k, \widehat{\lambda}_k) + \frac{\eta_{f,k}}{2\alpha_k^2} \|u - \widetilde{u}_k\|^2 \right\} \\ \bar{u}_{k+1} = u_{k+1} + (u_{k+1} - u_k)/\alpha_k \\ v_{k+1} = \mathbf{prox}_{\tau_k g}(\widetilde{v}_k - \tau_k B^* \bar{\lambda}_k) \\ \bar{v}_{k+1} = v_{k+1} + (v_{k+1} - v_k)/\alpha_k \\ \lambda_{k+1} = \lambda_k + \alpha_k/\theta_k(A\bar{u}_{k+1} + B\bar{v}_{k+1} - b) \end{cases} \tag{Fast-ADMM-1}$$

  with last iterate u_k, v_k, λ_k, intermediate sequence $\widehat{\lambda}_k, \bar{\lambda}_k, \widetilde{u}_k, \widetilde{v}_k$ and parameter sequence $\tau_k, \eta_{f,k}, \eta_{g,k}$.

- Sequential inner solvers: $\mathbf{prox}_{f+A^*A}$ and $\mathbf{prox}_g$

- Take the step size $\alpha_k^2 \|B\|^2 = \theta_k \beta_k$

- Take the step size $\alpha_k^2 \|B\|^2 = \theta_k \beta_k$
- Convergence rate

$$E_k + R_k \lesssim \min\left\{ \frac{\|B\|}{k}, \frac{\|B\|^2}{\mu_g k^2} \right\}$$

where

$$R_k := \underbrace{\mathcal{L}(u_k, v_k, \widehat{\lambda}) - \mathcal{L}(\widehat{u}, \widehat{v}, \lambda_k)}_{\text{Lagrange gap}}$$

$$E_k := \underbrace{\|Au_k + Bv_k - b\|}_{\text{Feasibility violation}} + \underbrace{\left|F(u_k, v_k) - \widehat{F}\right|}_{\text{Objective residual}}$$

- Take the step size $\alpha_k^2 \|B\|^2 = \theta_k \beta_k$
- Convergence rate

$$E_k + R_k \lesssim \min\left\{ \frac{\|B\|}{k}, \frac{\|B\|^2}{\mu_g k^2} \right\}$$

where

$$R_k := \underbrace{\mathcal{L}(u_k, v_k, \widehat{\lambda}) - \mathcal{L}(\widehat{u}, \widehat{v}, \lambda_k)}_{\text{Lagrange gap}}$$

$$E_k := \underbrace{\|Au_k + Bv_k - b\|}_{\text{Feasibility violation}} + \underbrace{\left| F(u_k, v_k) - \widehat{F} \right|}_{\text{Objective residual}}$$

- When $B = -I, b = 0$,

$$0 \leq P(u_k) - P(\widehat{u}) \lesssim \min\left\{ \frac{\|B\|}{k}, \frac{\|B\|^2}{\mu_g k^2} \right\}$$

- Take the step size $\alpha_k^2 \|B\|^2 = \theta_k \beta_k$
- Convergence rate

$$E_k + R_k \lesssim \min\left\{ \frac{\|B\|}{k}, \frac{\|B\|^2}{\mu_g k^2} \right\}$$

where

$$R_k := \underbrace{\mathcal{L}(u_k, v_k, \widehat{\lambda}) - \mathcal{L}(\widehat{u}, \widehat{v}, \lambda_k)}_{\text{Lagrange gap}}$$

$$E_k := \underbrace{\|Au_k + Bv_k - b\|}_{\text{Feasibility violation}} + \underbrace{\left|F(u_k, v_k) - \widehat{F}\right|}_{\text{Objective residual}}$$

- When $B = -I, b = 0$,

$$0 \le P(u_k) - P(\widehat{u}) \lesssim \min\left\{ \frac{\|B\|}{k}, \frac{\|B\|^2}{\mu_g k^2} \right\}$$

- Linearization of $f = f_1 + f_2$

$$\min\left\{ \frac{\|B\|}{\sqrt{\beta_0} k}, \frac{\|B\|^2}{\mu_g k^2} \right\} + \min\left\{ \frac{L_f}{\gamma_0 k^2}, \exp\left(-\frac{k}{4}\sqrt{\frac{\mu_f}{L_f}} \right) \right\}$$

More schemes

- The symmetric choice $\bar{\lambda}_{k+1} = \lambda_k + \alpha_k/\theta_k(A\bar{u}_k + B\bar{v}_{k+1} - b)$
  - Sequential inner solvers: $\mathbf{prox}_{g+B^*B}$ and $\mathbf{prox}_f$:

$$\min\left\{\frac{\|A\|}{k},\ \frac{\|A\|^2}{\mu_f k^2}\right\}$$

  - Linearization of $f = f_1 + f_2$

$$\min\left\{\frac{\|A\|}{\sqrt{\gamma_0}k} + \frac{L_f}{\gamma_0 k^2},\quad \frac{\|A\|^2}{\mu_f k^2} + \exp\left(-\frac{k}{4}\sqrt{\frac{\mu_f}{L_f}}\right)\right\}$$

More schemes

- The symmetric choice $\bar{\lambda}_{k+1} = \lambda_k + \alpha_k/\theta_k(A\bar{u}_k + B\bar{v}_{k+1} - b)$
  - Sequential inner solvers: $\mathbf{prox}_{g+B^*B}$ and $\mathbf{prox}_f$:

$$\min\left\{\frac{\|A\|}{k}, \frac{\|A\|^2}{\mu_f k^2}\right\}$$

  - Linearization of $f = f_1 + f_2$

$$\min\left\{\frac{\|A\|}{\sqrt{\gamma_0}k} + \frac{L_f}{\gamma_0 k^2}, \quad \frac{\|A\|^2}{\mu_f k^2} + \exp\left(-\frac{k}{4}\sqrt{\frac{\mu_f}{L_f}}\right)\right\}$$

- The second choice $\bar{\lambda}_{k+1} = \lambda_k + \alpha_k/\theta_k(A\bar{u}_k + B\bar{v}_k - b)$
  - Parallel inner solvers: $\mathbf{prox}_f$ and $\mathbf{prox}_g$

$$\min\left\{\frac{\|B\|}{k}, \frac{\|B\|^2}{\mu_g k^2}\right\} + \min\left\{\frac{\|A\|}{k}, \frac{\|A\|^2}{\mu_f k^2}\right\}$$

  - Linearization of $f = f_1 + f_2$

$$\min\left\{\frac{\|B\|}{\sqrt{\beta_0}k}, \frac{\|B\|^2}{\mu_g k^2}\right\} + \min\left\{\frac{\|A\|}{\sqrt{\gamma_0}k} + \frac{L_f}{\gamma_0 k^2}, \frac{\|A\|^2}{\mu_f k^2} + \exp\left(-\frac{k}{4}\sqrt{\frac{\mu_f}{L_f}}\right)\right\}.$$

- All can be extended to $g = g_1 + g_2$

LAD regression

Consider the least absolute deviation (LAD) regression problem

$$\min_{x \in \mathbb{R}^n} P(x) := f(x) + \|Ax - b\|_1 , \tag{4}$$

where $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$ are given data with $m \ll n$, and f is a regularization function. Consider two types of regularizer:

- *Case 1*: $f(x) = \lambda \|x\|_1$,
- *Case 2*: $f(x) = \lambda \|x\|_1 + \mu_f/2 \|x\|^2$.

We generate the matrix A from the standard normal distribution and set $b = Ax^{\#} + e$, where $x^{\#}$ is sparse and e is a Gaussian noise.

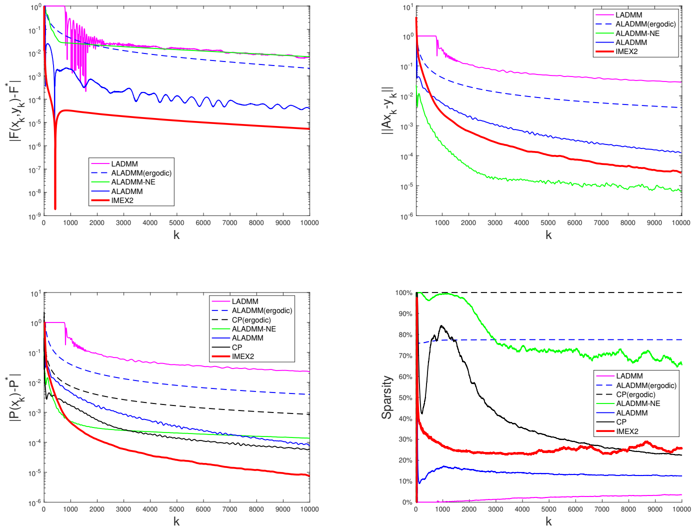

图 1: LAD regression, *Case 1*, with $(m, n) = (400, 4000)$ and sparsity 10%.

No worse (or comparable) convergence but good sparsity

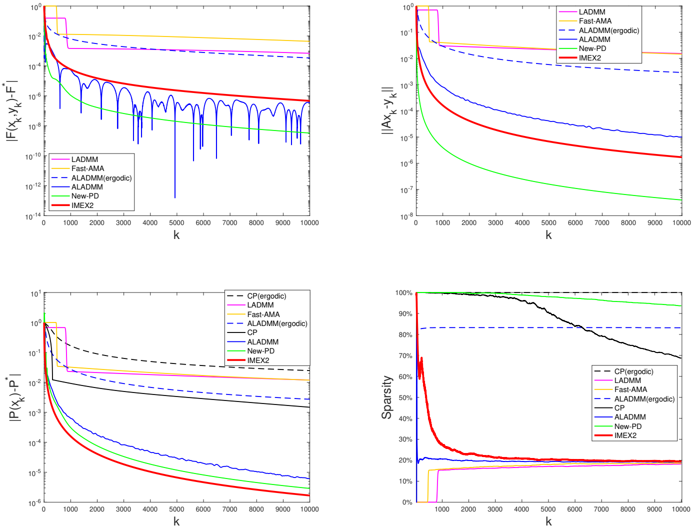

图 2: LAD regression, *Case 2*, with $(m, n) = (400, 4000)$ and sparsity 10%.

No worse (or comparable) convergence but good sparsity

References

This talk is mainly based on

Hao Luo, and ZiHang Zhang.
A unified differential equation solver approach for separable convex optimization: splitting, acceleration and nonergodic rate.
arXiv:2109.13467v2, 2023. (Under review)

Other related references

Hao Luo, and Long Chen.
From differential equation solvers to accelerated first-order methods for convex optimization. ***Math. Program.*** 195:735–781, 2022.

Hao Luo.
A universal accelerated primal-dual method for convex optimization problems. ***JOTA***, 2024.

Hao Luo.
A primal-dual flow for affine constrained convex optimization.
ESAIM : COCV, 2022.

Hao Luo.
Accelerated primal-dual methods for linearly constrained convex optimization problems. *arXiv:2109.12604*, 2021.

Thanks for your listening!

Any questions?