First-Order Methods in Convex Optimization: From Discrete to Continuous and Vice-versa

Hao Luo (罗浩)

National Center for Applied Mathematics in Chongqing
Chongqing Normal University

重庆师范大学 CHONGQING NORMAL UNIVERSITY
北京大學 PEKING UNIVERSITY 重庆大数据研究院 CHONGQING RESEARCH INSTITUTE OF BIG DATA

The 8-th Chinese–German Workshop on Computational and Applied Mathematics

School of Mathematics, Sichuan University
23th-27th Sept., 2024

Collaborators: Long Chen (UCI), Jun Hu (PKU),
ZiHang Zhang (PKU, PhD)

Outline

Introduction

From FOM to ODE

From ODE to FOM

Summary

Introduction

From FOM to ODE

From ODE to FOM

Summary

Problem setting

- Composite convex optimization (CCO) problem

$$\boxed{\inf_{x\in\mathbb{X}} F(x) := f(x) + g(Ax)} \tag{CCO}$$

Assumptions:
  - $\mathbb{X}$, $\mathbb{Y}$: Hilbert spaces with inner product $\langle\cdot,\cdot\rangle$ [1]
  - $A : \mathbb{X} \to \mathbb{Y}$: bounded linear operator
  - $f(g) : \mathbb{X}(\mathbb{Y}) \to (-\infty, +\infty]$: CCP [2] with constants $\mu_f(\mu_g) \geq 0$
  - Consistent condition: $A\mathbf{dom}\, f \cap \mathbf{dom}\, g \neq \emptyset$
- Linearly constrained optimization (LCO) problem

$$\boxed{\inf_{x\in\mathbb{X}} f(x) \quad \text{s.t. } Ax = b} \tag{LCO}$$

- Bilinear saddle-point (BSP) problem

$$\boxed{\inf_{x\in\mathbb{X}} \sup_{y\in\mathbb{Y}} \mathcal{L}(x,y) := f(x) + \langle y, Ax\rangle - g(y)} \tag{BSP}$$

- Many applications in:
  - TV model (Image processing), Machine learning ...
  - p-Laplacian (Numerical PDEs), Optimal transport, ...

[1] When no confusion arises, we use the same bracket $\langle\cdot,\cdot\rangle$ for the inner products on $\mathbb{X}$ and $\mathbb{Y}$.

[2] CCP means closed, convex and proper.

Optimality condition and Algorithm class

- First-order optimality conditions:

$$\text{For (CCO)} \quad 0 \in \partial f(x^*) + A^\star \partial g(Ax^*)$$
$$\text{For (LCO)} \quad 0 \in \begin{bmatrix} \partial f(x^*) + A^\star y^* \\ b - Ax^* \end{bmatrix}$$
$$\text{For (BSP)} \quad 0 \in \begin{bmatrix} \partial f(x^*) + A^\star y^* \\ \partial g^\star(y^*) - Ax^* \end{bmatrix}$$

- A unified abstract presentation: Finding a zero point $0 \in M(\mathbf{x}^*)$ of a maximal monotone operator $M : \mathcal{X} \to 2^{\mathcal{X}}$.
- We are mainly interested in First-Order Methods (FOM) that produce the iteration sequence $\{x_k\}$ with the access **only** to[3]

$$\nabla f/\mathbf{prox}_f, \quad \nabla g/\mathbf{prox}_g$$

or (for $f = f_1 + f_2,\ g = g_1 + g_2$)

$$\nabla f_1/\mathbf{prox}_{f_2}, \quad \nabla g_1/\mathbf{prox}_{g_2}$$

[3] Here and in what follows, $\mathbf{prox}_f$ denotes the **proximal mapping** of f:

$$\mathbf{prox}_f(x) = \operatorname{argmin}\left\{f(y) + 1/2\,\|x - y\|^2\right\}$$

Proximal-gradient methods for (CCO) with $A = I$

- Gradient descent (GD) and Proximal point algorithm (PPA):

$$x_{k+1} = x_k - s\nabla F(x_k), \quad x_{k+1} = x_k - s\nabla F(x_{k+1})^4$$

- Proximal-gradient method (PGM): $x_{k+1} = x_k - s(\nabla f(x_k) + \nabla g(x_{k+1}))$
  - Also known as Forward-Backward Splitting
  - $O(1/k)$ for convex and $(1 - 1/\kappa_f)$ for strongly convex

- Heavy ball (HB)[5]: $x_{k+1} = x_k - s\nabla F(x_k) + \beta_k \overbrace{(x_k - x_{k-1})}^{\text{Momentum}}$
  - Better than GD with $\beta_k \in (0, 1)$
  - Optimal choice of strongly convex case

- Nesterov accelerated gradient (NAG-1983, NAG-2004):

$$x_{k+1} = \bar{x}_k - s\nabla F(\bar{x}_k), \quad \bar{x}_{k+1} = x_{k+1} + \beta(x_{k+1} - x_k)$$

  - $O(1/k^2)$ with $\beta_k = k/(k+3)$
  - $O(1 - 1/\sqrt{\kappa_f})^k$ with $\beta_k = (\sqrt{\kappa_f} - 1)/(\sqrt{\kappa_f} + 1)$
  - Optimal rate
  - Proximal gradient version = FISTA

- Güler's PPA (**SIOPT**, 1994)

$$x_{k+1} = \bar{x}_k - s\nabla F(x_{k+1}), \quad \bar{x}_{k+1} = x_{k+1} + \beta(x_{k+1} - x_k)$$

---

[4] This presentation is equivalent to $x_{k+1} = \mathbf{prox}_{sF}(x_k)$

[5] Polyak, 1964

Augmented Lagrangian methods for (LCO)

- Augmented Lagrangian method (ALM)

$$x_{k+1} = \operatorname*{argmin}_{x\in\mathbb{X}} \left\{ \mathcal{L}(x, \lambda_k) + \frac{\sigma}{2} \|Ax - b\|^2 \right\}, \quad \lambda_{k+1} = \lambda_k + \sigma(Ax_{k+1} - b)$$

- Equivalent to Bregman method and dual PPA
- Linearization (L-ALM) and relaxation (ADMM)
- $O(1/k^2)$ acceleration with momentum for the dual variable [6]
- Acceleration with momentum for the primal variable [7]
    * $O(\frac{1}{k})$ for convex and $O(\frac{1}{k^2})$ for strongly convex (Optimal) [8]
    * Extension to two block case (Acc-ADMM) [9]

[6] He and Yuan, 2013; Kang et al. **JSC**, 2013

[7] Xu, **SIOPT**, 2017

[8] Ouyang and Xu, **SIOPT**, 2021

[9] Sabach and Teboulle, **SIOPT**, 2022; Zhang et al. arXiv:2206.05088, 2022

Primal-dual methods for (BSP)

- Extensions of GD and PPA:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - sM(\mathbf{x}_k), \qquad \mathbf{x}_{k+1} = \mathbf{x}_k - sM(\mathbf{x}_{k+1})$$

Diverge — Full coupling

- Extra-gradient method (EGM, with ergodic rate $O(1/k)$) [10]

$$\bar{\mathbf{x}}_k = \mathbf{x}_k - sM(\mathbf{x}_k), \quad \mathbf{x}_{k+1} = \mathbf{x}_k - sM(\bar{\mathbf{x}}_k)$$

- Primal-dual hybrid gradient method (PDHG) (Preconditioned PPA)

$$\mathbf{x}_{k+1} = \mathbf{x}_k - sQ^{-1}M(\mathbf{x}_{k+1}), \quad Q = \begin{bmatrix} I & -sA^\star \\ O & I \end{bmatrix}$$

- Also known as the primal-dual proximal splitting (PDPS)

$$\begin{cases} x_{k+1} = \mathbf{prox}_{sf}(x_k - sA^\star y_k) \\ y_{k+1} = \mathbf{prox}_{sg}(y_k + sAx_{k+1}) \end{cases}$$

- Diverge even for LP (He et al. **JMIV**, 2017)

[10] Ergodic means for the average $\bar{\mathbf{x}}_N = \sum_{i=0}^N a_i \mathbf{x}_i / \sum_{i=0}^N a_i$

- A symmetrized precondition remedy

$$\mathbf{x}_{k+1} = \mathbf{x}_k - sQ^{-1}M(\mathbf{x}_{k+1}), \quad Q = \begin{bmatrix} I & -sA^\star \\ -sA & I \end{bmatrix}$$

- This is the Chambolle–Pock (CP) [11]

$$\begin{cases} x_{k+1} = \mathbf{prox}_{sf}(x_k - sA^\star y_k) \\ y_{k+1} = \mathbf{prox}_{sg}(y_k + sA(2x_{k+1} - x_k)) \end{cases}$$

- Optimal ergodic rate: $O(1/k)$ for convex , $O(1/k^2)$ for partially strongly convex and ρ^k for strongly convex
- Inertial corrected PDPS [12] (IC-PDPS, with momentum and correction)

$$\begin{cases} \mathbf{x}_{k+1} = \bar{\mathbf{x}}_k - Q_{k+1}^{-1}M(\mathbf{x}_{k+1}) + \underbrace{\widehat{Q}_{k+1}(\mathbf{x}_{k+1} - \mathbf{x}_k)}_{\text{Correction}}, \\ \bar{\mathbf{x}}_{k+1} = \mathbf{x}_{k+1} + \Lambda_{k+1}(\mathbf{x}_{k+1} - \mathbf{x}_k), \end{cases}$$

- Optimal nonergodic rate

[11] Chambolle and Pock, **JMIV**, 2013

[12] Valkonen, **SIOPT**, 2020

Motivation

- Almost all FOMs (without momentum) in the form

$$X^+ = \Gamma(s, X)$$

- This is very close to **Numerical Discretization**
- Can we have a unified continuous perspective on FOMs?
- How about the numerical analysis approach for FOMs?

Introduction

From FOM to ODE

From ODE to FOM

Summary

$O(s^r)$-resolution framework

Definition 1 (Lu, MAPR, 2022)

Given a FOM $X^+ = \Gamma(s, X)$ with $\Gamma(0, X) = X$, if there is an ODE system

$$X' = \Gamma_0(X) + s\Gamma_1(X) + \cdots + s^r\Gamma(X) \tag{1}$$

that satisfies $\left\| X(s) - X^+ \right\| = o\left(s^{r+1}\right)$ with $r \geq 0$, where $X(s)$ is the solution of (1) with $X(0) = X$, then we call (1) the $O(s^r)$-resolution ODE of the FOM $X^+ = \Gamma(s, X)$

Theorem 1 (Lu, MAPR, 2022)

Given a FOM $X^+ = \Gamma(s, X)$ with $\Gamma(0, X) = X$ and sufficiently smooth $\Gamma(s, X)$ in both s and X Then its $O(s^r)$-resolution ODE exists uniquely.

$O(s^r)$-resolution without momentum

Look at $E(s) = X(s) - X^+ = X - \Gamma(s, X) + \int_0^s X'(t, s)\,\mathrm{d}t$ and the Taylor expansion at $s = 0$

$$E(s) = E(0) + E'(0)s + \cdots + \frac{E^{(r+1)}}{(r+1)!}s^{r+1} + o(s^{r+1})$$

Essentially, we have $E(0) = E'(0) = \cdots = E^{(j)}(0) = 0$. This gives Γ_j

Corollary 1 (Lu, **MAPR**, 2022)

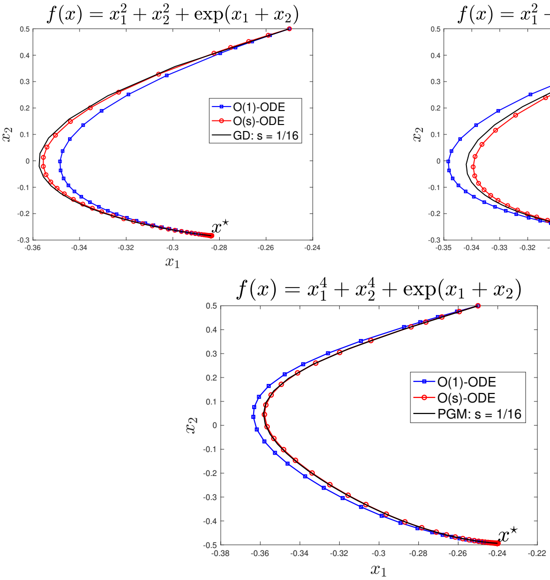
(i) *The $O(1)$-resolution ODE of GD, PPA and PGM:* $X' = -\nabla F(X)$

(ii) *The $O(s)$-resolution ODE of GD is* $X' = -\nabla F(X) - \frac{s}{2}\nabla^2 F(X) \cdot \nabla F(X)$

(iii) *The $O(s)$-resolution ODE of PPA is* $X' = -\nabla F(X) + \frac{s}{2}\nabla^2 F(X) \cdot \nabla F(X)$

(iv) *The $O(s)$-resolution ODE of PGM is*

$$X' = -\nabla F(X) + \frac{s}{2}(\nabla^2 g(X) - \nabla^2 f(X)) \cdot \nabla F(X)$$



$f(x) = x_1^2 + x_2^2 + \exp(x_1 + x_2)$
O(1)-ODE
O(s)-ODE
GD: s = 1/16
$x^\star$
x_1
x_2
$f(x) = x_1^2 + x_2^2 + \exp(x_1 + x_2)$
O(1)-ODE
O(s)-ODE
PPA: s = 1/16
$x^\star$
x_1
x_2
$f(x) = x_1^4 + x_2^4 + \exp(x_1 + x_2)$
O(1)-ODE
O(s)-ODE
PGM: s = 1/16
$x^\star$
x_1
x_2

Corollary 2 (Lu, **MAPR**, 2022)

(i) *The $O(1)$-resolution ODE of GD, PPA, PDHG, CP and EGM are*

$$X' = -M(X)$$

(ii) *The $O(s)$-resolution ODE of GD is*

$$X' = -M(X) - \frac{s}{2}\nabla M(X) \cdot M(X)$$

(iii) *The $O(s)$-resolution ODE of PPA and EGM are the same*

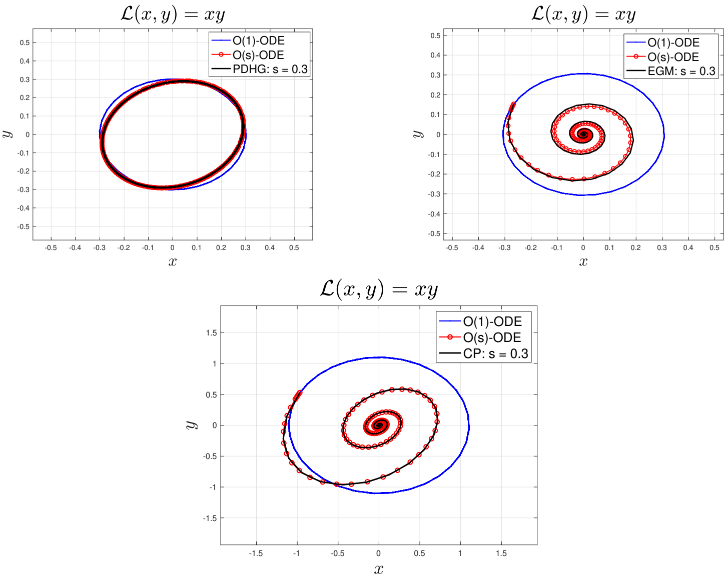
$$X' = -M(X) + \frac{s}{2}\nabla M(X) \cdot M(X)$$

(iv) *The $O(s)$-resolution ODE of PDHG is*

$$X' = -M(X) + \frac{s}{2}\left[\nabla M(X) + \begin{bmatrix} O & O \\ 2A & O \end{bmatrix}\right] \cdot M(X)$$

(iv) *The $O(s)$-resolution ODE of CP is*

$$X' = -M(X) + \frac{s}{2}\left[\nabla M(X) + \begin{bmatrix} O & 2A^{\star} \\ 2A & O \end{bmatrix}\right] \cdot M(X)$$



$\mathcal{L}(x,y) = xy$
O(1)-ODE
O(s)-ODE
PDHG: s = 0.3
x
y
$\mathcal{L}(x,y) = xy$
O(1)-ODE
O(s)-ODE
EGM: s = 0.3
x
y
$\mathcal{L}(x,y) = xy$
O(1)-ODE
O(s)-ODE
CP: s = 0.3
x
y

$O(s^r)$-resolution with momentum

- For a general momentum method

$$x_{k+1} = x_k - s\nabla F(x_k) + \underbrace{\beta(s)(x_k - x_{k-1})}_{\text{Momentum}} - \beta(s)s\left[\nabla F(x_k) - \nabla F(x_{k-1})\right]$$

  there is No such condition $\Gamma(0, X) = 0$.

- Key observation: A hybrid gradient descent transformation

$$\frac{x_{k+1} - x_k + s\nabla F(x_k)}{\sqrt{s}\beta(s)} = \beta(s) \cdot \frac{x_k - x_{k-1} + s\nabla F(x_{k-1})}{\sqrt{s}\beta(s)} - \sqrt{s}\nabla F(x_k)$$

  which leads to

$$\begin{cases} x_{k+1} = x_k + \sqrt{s}\beta(s)v_{k+1} - s\nabla F(x) \\ v_{k+1} = v_k + (\beta(s) - 1)v_k - \sqrt{s}\nabla F(x) \end{cases}$$

  with $\lim\limits_{s\to 0}(\beta(s) - 1)/\sqrt{s} = 0$

- This gives a new system of $X = (x, v)$ that satisfies $X^+ = \Gamma(\sqrt{s}, X)$ with $\Gamma(0, X) = 0$
- The same idea works for other momentum methods with dynamically changing parameters and primal-dual methods

Theorem 2

(i) *The $O(1)$-resolution ODE of HB and NAG with optimal β for strongly convex objective are the same* [13]*:*

$$\begin{bmatrix} x \\ v \end{bmatrix}' = \begin{bmatrix} v \\ -2\sqrt{\mu}v - \nabla F(x) \end{bmatrix} \quad \Longleftrightarrow \quad x'' + 2\sqrt{\mu}x' + \nabla F(x) = 0$$

(ii) *The $O(1)$-resolution ODE of NAG-1983/FISTA for convex objective is*

$$\begin{bmatrix} x \\ v \\ \gamma \end{bmatrix}' = \begin{bmatrix} v \\ -\frac{3}{2\sqrt{\gamma}}v - \nabla F(x) \\ \sqrt{\gamma} \end{bmatrix} \quad \Longleftrightarrow \quad x'' + \frac{3}{2\sqrt{\gamma}}x' + \nabla F(x) = 0$$

*Since $\gamma = t^2/4$, this gives the Su-Boyd-Candès (**JMLR**, 2016)*

$$x'' + \frac{3}{t}x' + \nabla F(x) = 0$$

[13] Polyak. 1964; Siegel. 2019; Wilson et al. ***JMLR***, 2021; Shi et al., ***Math. Program.***, 2022;

(iii) The $O(1)$-resolution ODE of NAG-2004 is [14]

$$\begin{bmatrix} x \\ v \\ \gamma \end{bmatrix}' = \begin{bmatrix} v \\ -\frac{3+\mu\gamma}{2\sqrt{\gamma}} v - \nabla F(x) \\ \sqrt{\gamma}(1-\mu\gamma) \end{bmatrix} \quad \Longleftrightarrow \quad x'' + \frac{3+\mu\gamma}{2\sqrt{\gamma}} x' + \nabla F(x) = 0$$

(iv) The $O(1)$-resolution ODE of IC-PDPS is [15]

$$\begin{bmatrix} \mathbf{x} \\ \mathbf{v} \\ \Upsilon \\ \theta \end{bmatrix}' = \begin{bmatrix} \mathbf{v} - \mathbf{x} \\ -\theta\Upsilon^{-1} \left[S(\mathbf{v} - \mathbf{x}) + M(\mathbf{x}) \right] \\ 2\mathrm{diag}(S)\Upsilon \\ \theta \end{bmatrix}$$

In second-order form

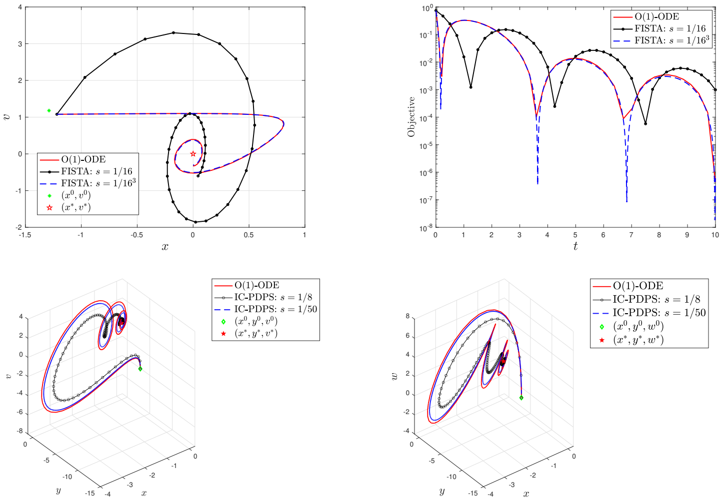
$$\Upsilon \mathbf{x}'' + [\theta S + \Upsilon]\, \mathbf{x}' + \theta M(\mathbf{x}) = 0, \quad S = \begin{bmatrix} \mu_f I & A^\star \\ A & \mu_g I \end{bmatrix}$$

In component wise

$$\begin{cases} \gamma x'' + (\gamma + \mu_f \theta) x' + \theta \nabla_x \mathcal{L}(x, y + y') = 0 \\ \beta y'' + (\beta + \mu_g \theta) y' + \theta \nabla_y \mathcal{L}(x + x', y) = 0 \end{cases}$$

[14] L., and Long Chen. ***Math. Program.***, 2022.

[15] L. *arXiv:2405.14098v1*, 2024.



O(1)-ODE
FISTA: $s = 1/16$
FISTA: $s = 1/16^3$
(x^0, v^0)
(x^*, v^*)
x
v
Objective
t
IC-PDPS: $s = 1/8$
IC-PDPS: $s = 1/50$
(x^0, y^0, v^0)
(x^*, y^*, v^*)
(x^0, y^0, w^0)
(x^*, y^*, w^*)
y
w

Introduction

From FOM to ODE

From ODE to FOM

Summary

Semi-implicit AGD

For unconstrained minimization problem, we present a compact form of the $O(1)$-resolution ODE of NAG-2004 with time scaling:

$$\begin{aligned} \gamma x'' + (\mu + \gamma) x' + \nabla F(x) &= 0 \\ \gamma' - \mu + \gamma &= 0 \end{aligned} \qquad \text{(NAG flow)}$$

- Semi-implicit scheme for Accelerated Gradient Descent (AGD) [16]

$$\gamma_k \cdot \frac{\frac{x_{k+1}-x_k}{\alpha_k} - \frac{x_k - x_{k-1}}{\alpha_{k-1}}}{\alpha_k} + (\mu + \gamma_k) \cdot \frac{x_{k+1} - x_k}{\alpha_k} + \nabla F(\bar{x}_k) = 0$$

- Composite case $F = f + g$

$$\gamma_k \cdot \frac{\frac{x_{k+1}-x_k}{\alpha_k} - \frac{x_k - x_{k-1}}{\alpha_{k-1}}}{\alpha_k} + (\mu + \gamma_k) \cdot \frac{x_{k+1} - x_k}{\alpha_k} + \nabla f(\bar{x}_k) + \nabla g(x_{k+1}) = 0$$

- Lyapunov analysis (optimal rate)

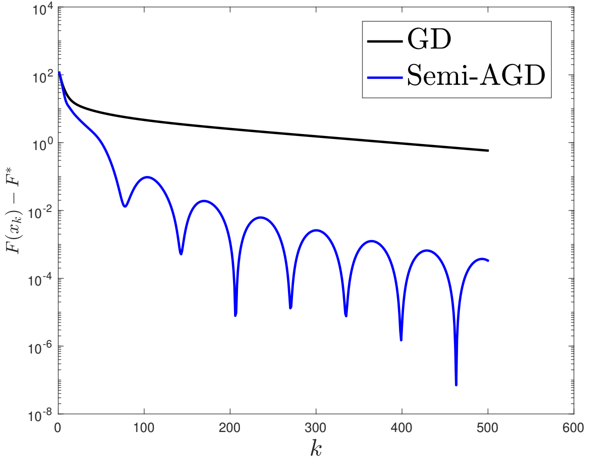
$$\mathcal{E}_k := F(x_k) - F(x^*) + \frac{\gamma_k}{2} \left\| v_k - x^* \right\|^2 \le \min \left\{ \frac{L}{k^2}, \left(1 + \sqrt{\frac{\mu_f}{L_f}} \right)^{-k} \right\}$$

[16] L., and Long Chen. ***Math. Program.***, 2022/*arXiv:1912.09276*, 2019; L. ***Optimization***, 2023.

Find $u \in H_0^1(\Omega)$ such that

$$-\Delta u = f \quad \text{in } \Omega = (0,1)^2$$

Use $P1$ Lagrange element with uniform mesh size $h = 2^{-5}$. The DoF is $N = \dim V_h = (1/h+1)^2 = 1089$.



Restarting

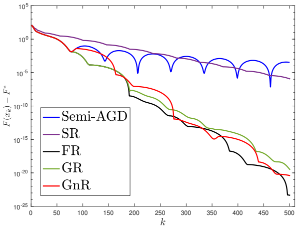
- Restarting scheme

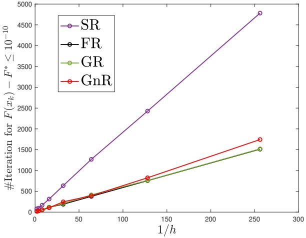
Function restart (FR) : $\dfrac{\mathrm{d}F(x(t))}{\mathrm{d}t} > 0$ O'Donoghue and Candès(**FoCM**, 2015)

Gradient restart (GR) : $\left\langle \nabla F(x(t)), x'(t) \right\rangle > 0$ O'Donoghue and Candès,(**FoCM**, 2015)

Speed restart (SR) : $\dfrac{\mathrm{d}\left\|x'(t)\right\|^2}{\mathrm{d}t} < 0$ Su-Boyd-Candès (***JMLR***, 2016)

Gradient norm restart (GnR) : $\dfrac{\mathrm{d}\left\|\nabla F(x(t))\right\|^2}{\mathrm{d}t} > 0$





Restart works very well with the iteration complexity $\sim \sqrt{\kappa}$

This yield the linear rate $\exp\left(-k/\sqrt{\kappa}\right)$

Implicit-explicit AALM

For (LCO), we propose a simplified form of the $O(1)$-resolution ODE of IC-PDPS:

$$
\begin{aligned}
\gamma x'' + (\mu+\gamma)x' + \nabla f(x) + A^\top y &= 0\\
\beta y' + b - A(x+x') &= 0\\
\gamma' - \mu + \gamma &= 0\\
\beta' + \beta &= 0
\end{aligned}
\qquad \text{(APD flow)}
$$

- Implicit-explicit scheme for Accelerated Augmented Lagrangian Method (AALM) [17]

$$
\begin{aligned}
\gamma_k \cdot \frac{\frac{x_{k+1}-x_k}{\alpha_k} - \frac{x_k - x_{k-1}}{\alpha_{k-1}}}{\alpha_k} + (\mu+\gamma_k)\cdot \frac{x_{k+1}-x_k}{\alpha_k} + \nabla f(\bar{x}_k) + A^\top \bar{y}_k &= 0\\
\beta_k \frac{y_{k+1}-y_k}{\alpha_k} + b - A\left(x_{k+1} + (x_{k+1}-x_k)/\alpha_k\right) &= 0
\end{aligned}
$$

- Lyapunov analysis (optimal nonergodic rate)

$$
\mathcal{E}_k := \mathcal{L}(x_k, y^*) - \mathcal{L}(x^*, y_k) + \frac{\gamma_k}{2}\left\|v_k - x^*\right\|^2 + \frac{\beta_k}{2}\left\|y_k - y^*\right\|^2 \le \begin{cases} Ck^{-1}, & \mu = 0,\\ Ck^{-2}, & \mu > 0. \end{cases}
$$

[17] L. *arXiv:2109.12604v2*, 2023.

- For extension to the Hölder case $\nabla f \in C^{0,\nu}$ and application to optimal transport (ODE+AMG+SsN), see Hu et al. (JSC,2023) and L. (JOTA, 2024).

- For the separable case $f(x) = f_1(x_1) + f_2(x_2)$, we have implicit-explicit schemes for accelerated ADMM; see L. and Zhang (arXiv:2109.13467v2, 2023).

Semi-implicit APDGS

For (BSP), we have a simplified form of the $O(1)$-resolution ODE of IC-PDPS:

$$\begin{aligned} \Upsilon \mathbf{x}'' + [S + \Upsilon]\,\mathbf{x}' + M(\mathbf{x}) &= 0 \\ \Upsilon' - \Sigma + \Upsilon &= 0 \end{aligned} \qquad \text{(APDG flow)}$$

- Implicit-explicit scheme for Accelerated Primal-Dual Gradient Splitting (APDGS) [18]

$$\gamma_k \cdot \frac{\frac{x_{k+1}-x_k}{\alpha_k} - \frac{x_k - x_{k-1}}{\alpha_{k-1}}}{\alpha_k} + (\mu_f + \gamma_k)\cdot \frac{x_{k+1} - x_k}{\alpha_k} + \nabla f(\bar{x}_k) + A^\top \bar{y}_k = 0$$

$$\beta_k \cdot \frac{\frac{y_{k+1}-y_k}{\eta_k \alpha_k} - \frac{y_k - y_{k-1}}{\eta_{k-1}\alpha_{k-1}}}{\alpha_k} + (\mu_g/\eta_k + \beta_k)\cdot \frac{y_{k+1} - y_k}{\alpha_k} + \eta_k(\nabla g(\bar{y}_k) - A\bar{x}_{k+1}) = 0$$

- Lyapunov analysis (optimal nonergodic rate)

$$\mathcal{E}_k = \mathcal{L}(x_k, y^*) - \mathcal{L}(x^*, y_k) + \frac{\gamma_k}{2}\left\| v_k - x^* \right\|^2 + \frac{\beta_k}{2}\left\| w_k - y^* \right\|^2 \le \begin{cases} \frac{C}{k}, & \mu_f = \mu_g = 0, \\ \frac{C}{k^2}, & \mu_f + \mu_g > 0, \\ \rho^k, & \mu_f \mu_g > 0, \end{cases}$$

[18] L. *arXiv:2407.20195*, 2024.

Introduction

From FOM to ODE

From ODE to FOM

Summary

Summary

- Conclusion:
  - * A unified $O(s^r)$-resolution framework for FOMs
  - * A time discretization approach to construct FOMs
  - * A Lyapunov function analysis for optimal convergence rate
  - * Some numerical illustration with restarting

- Future topics:
  - * Extension to nonlinear saddle-point problems (General convex optimization with nonlinear but convex constraint)
  - * Restarting with uniform convergence rate independent on the condition number (Multilevel + restarting)
  - * Restart analysis for the primal-dual dynamics (No descent)
  - * Application to nonlinear variational problems (Nonconvex but with special structure) and optimal transport
  - * Accelerated multiobjective gradient methods

References

Hao Luo.

A universal accelerated primal-dual method for convex optimization problems. ***J.Optim.Theory Appl.***, 201(1):280-312, 2024.

Hao Luo, and ZiHang Zhang.

A unified differential equation solver approach for separable convex optimization: splitting, acceleration and nonergodic rate. *arXiv:2109.13467v2*, 2023. (Submitted to ***Math. Comp.*** Under review)

Jun Hu, Hao Luo, and ZiHang Zhang.

A fast solver for generalized optimal transport problems based on dynamical system and algebraic multigrid. ***J.Sci.Comput.***, 97(6): https://doi.org/10.1007/s10915-023-02272-9, 2023.

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 primal-dual flow for affine constrained convex optimization. ***ESAIM Control Optim.Calc.Var.***, 28:33, 2022.

References

Hao Luo.
A continuous perspective on the inertial corrected primal-dual proximal splitting. *arXiv:2405.14098v1*, 2024.

Hao Luo.
Accelerated primal-dual proximal gradient splitting methods for convex-concave saddle-point problems. *arXiv:2407.20195*, 2024.

Hao Luo.
Accelerated differential inclusion for convex optimization. ***Optimization***, 72(5):1139–1170, 2023.

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

Hao Luo, and Long Chen.
First order optimization methods based on Hessian-driven Nesterov accelerated gradient flow. *arXiv:1912.09276*, 2019.

Thanks for your listening!

Any questions?