A Fast First-Order Optimization Approach to Quasistatic Elastoplastic Analysis with von Mises Yield Criterion

Wataru Shimizu† Yoshihiro Kanno‡

†The University of Tokyo (IHI Corporation)

‡Tokyo Institute of Technology

September 5, 2017 (COMPLAS 2017)

theme

- equilibrium analysis w/ von Mises criterion $\simeq$ group LASSO
  - optimization approach to computational plasticity
    - min. potential energy [Maier '68, '70], etc.
  - data science
    - needs for solving large-scale (convex) optimization
    - least squares w/ regularization
      - LASSO [Tibshirani '96] group LASSO [Yuan & Lin '06]
  - fast 1st-order algo. for (convex) optim.
    - accelerated gradient method [Nesterov '83]
    - accelerated proximal grad. meth. [Beck & Teboulle '09]
  - Y. K.: "A fast first-order optimization approach to elastoplastic analysis of skeletal structures." *Optim. Eng.*, **17**, 861–896 (2016).

theme

- equilibrium analysis w/ von Mises criterion $\simeq$ group LASSO
  - optimization approach to computational plasticity
    - min. potential energy **(!)** [Maier '68, '70], etc.
  - data science
    - needs for solving large-scale (convex) optimization
    - least squares w/ regularization
      - LASSO [Tibshirani '96] group LASSO [Yuan & Lin '06] **(!)**
  - fast 1st-order algo. for (convex) optim.
    - accelerated gradient method [Nesterov '83]
    - accelerated proximal grad. meth. **(!)** [Beck & Teboulle '09]
  - Y. K.: "A fast first-order optimization approach to elastoplastic analysis of skeletal structures." *Optim. Eng.*, **17**, 861–896 (2016).

(accelerated) proximal gradient method

- 1st-order meth. for convex optim. [Bruck '77], [Passty '79], etc.
  - generalization of projected gradient method
  - can solve nonsmooth problems

(accelerated) proximal gradient method

- 1st-order meth. for convex optim. [Bruck '77], [Passty '79], etc.
  - generalization of projected gradient method
  - can solve nonsmooth problems
- application to ℓ_1-regularized least squares (LASSO)
  - ISTA (iterative shrinkage-thresholding algorithms)
    [Chambolle, DeVore, Lee, & Lucier '98], [Daubechies, Defrise, & Mol '04]
    - $O(1/k)$ convergence in obj. value (k: iteration counter)
  - acceleration: FISTA (fast ISTA) [Beck & Teboulle '09]
    - $O(1/k^2)$
  - restart of acceleration [O'Donoghue & Candes '15]
    - monotonic decrease of obj. value

proximal gradient method

- convex optimization:

$$\text{Minimize} \quad f(\boldsymbol{x}) + g(\boldsymbol{x})$$

  - f : convex, differentiable g : convex

proximal gradient method

- convex optimization:

$$\text{Minimize} \quad f(\boldsymbol{x}) + g(\boldsymbol{x})$$

  - f : convex, differentiable g : convex

- iteration:

$$\boldsymbol{x}^{(k+1)} := \mathbf{prox}_{\alpha g}(\boldsymbol{x}^{(k)} - \alpha \nabla f(\boldsymbol{x}^{(k)}))$$

  - def. of proximal mapping:

$$\mathbf{prox}_{\alpha g}(\boldsymbol{x}) = \arg\min_{\boldsymbol{z}} \left\{ \alpha g(\boldsymbol{z}) + \frac{1}{2}\|\boldsymbol{z} - \boldsymbol{x}\|^2 \right\}$$

  - $\alpha \in (0, 1/L]$: step size (L : Lipschitz constant of ∇f)
  - useful if computation of $\mathbf{prox}_{\alpha g}$ is easy
    $\simeq$ if g has a simple form

accelerated proximal gradient method

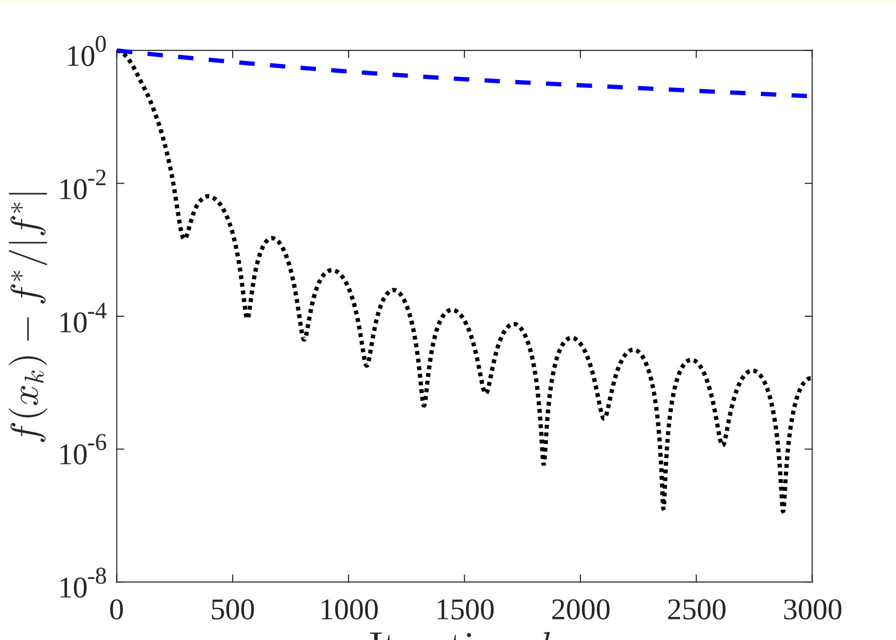
- original version — $O(1/k)$:

$$\boldsymbol{x}^{(k+1)} := \mathbf{prox}_{\alpha g}(\boldsymbol{x}^{(k)} - \alpha \nabla f(\boldsymbol{x}^{(k)}))$$

- accelerate version — $O(1/k^2)$: [Beck & Teboulle '09]

$$\boldsymbol{x}^{(k+1)} := \mathbf{prox}_{\alpha g}(\boldsymbol{y}^{(k)} - \alpha \nabla f(\boldsymbol{y}^{(k)}))$$
$$\boldsymbol{y}^{(k+1)} := \boldsymbol{x}^{(k+1)} + \omega^{(k)}(\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)})$$

  - e.g., $\omega^{(k)} := k/(k+3)$



accelerated proximal gradient method

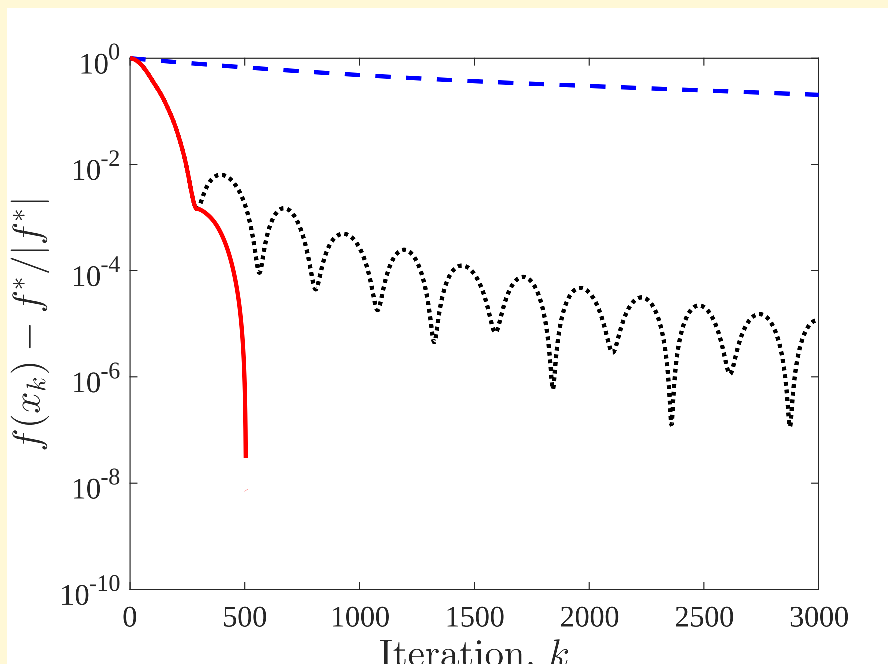
- original version — $O(1/k)$:

$$\boldsymbol{x}^{(k+1)} := \mathbf{prox}_{\alpha g}(\boldsymbol{x}^{(k)} - \alpha \nabla f(\boldsymbol{x}^{(k)}))$$

- accelerate version — $O(1/k^2)$: [Beck & Teboulle '09]

$$\boldsymbol{x}^{(k+1)} := \mathbf{prox}_{\alpha g}(\boldsymbol{y}^{(k)} - \alpha \nabla f(\boldsymbol{y}^{(k)}))$$
$$\boldsymbol{y}^{(k+1)} := \boldsymbol{x}^{(k+1)} + \omega^{(k)}(\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)})$$

- restart (reset $\omega^{(k)} := 0$) — monotonicity [O'Donoghue & Candes '15]



accelerated proximal gradient method

- original version — $O(1/k)$:

$$\boldsymbol{x}^{(k+1)} := \textbf{prox}_{\alpha g}(\boldsymbol{x}^{(k)} - \alpha \nabla f(\boldsymbol{x}^{(k)}))$$

- accelerate version — $O(1/k^2)$: [Beck & Teboulle '09]

$$\boldsymbol{x}^{(k+1)} := \textbf{prox}_{\alpha g}(\boldsymbol{y}^{(k)} - \alpha \nabla f(\boldsymbol{y}^{(k)}))$$
$$\boldsymbol{y}^{(k+1)} := \boldsymbol{x}^{(k+1)} + \omega^{(k)}(\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)})$$

- restart (reset $\omega^{(k)} := 0$) — monotonicity [O'Donoghue & Candes '15]
  - easy to implement
  - fast convergence
  - applicable to large-scale problems
    - most of computation: matrix-vector products
      no system of linear eqs.

accelerated proximal gradient method

- original version — $O(1/k)$:

$$\boldsymbol{x}^{(k+1)} := \mathbf{prox}_{\alpha g}(\boldsymbol{x}^{(k)} - \alpha \nabla f(\boldsymbol{x}^{(k)}))$$

- accelerate version — $O(1/k^2)$: [Beck & Teboulle '09]

$$\boldsymbol{x}^{(k+1)} := \mathbf{prox}_{\alpha g}(\boldsymbol{y}^{(k)} - \alpha \nabla f(\boldsymbol{y}^{(k)}))$$
$$\boldsymbol{y}^{(k+1)} := \boldsymbol{x}^{(k+1)} + \omega^{(k)}(\boldsymbol{x}^{(k+1)} - \boldsymbol{x}^{(k)})$$

- restart (reset $\omega^{(k)} := 0$) — monotonicity [O'Donoghue & Candes '15]
- "Unfortunately, it is very difficult to obtain strong intuition about the mechanism by which this remarkable phenomenon occurs."
  D. P. Bertsekas: *Convex Optimization Algorithms*, p. 322.

LASSO (least absolute shrinkage and selection operator)

- solves regularized least squares: [Tibshirani '96]

$$\text{Minimize} \quad \frac{1}{2}\|A\boldsymbol{x} - \boldsymbol{b}\|_2^2 + \gamma \sum_{j=1}^{n} |x_j|$$

$\gamma > 0$: parameter

- to find a sparse solution to a linear regression problem

LASSO (least absolute shrinkage and selection operator)

- solves regularized least squares: [Tibshirani '96]

$$\text{Minimize} \quad \underbrace{\frac{1}{2}\|A\boldsymbol{x} - \boldsymbol{b}\|_2^2}_{f(\boldsymbol{x})} + \underbrace{\gamma \sum_{j=1}^{n} |x_j|}_{g(\boldsymbol{x})}$$

$\gamma > 0$: parameter

  - to find a sparse solution to a linear regression problem

- ISTA = a proximal gradient method for LASSO:

$$\boldsymbol{x}^{(k+1)} = \mathbf{prox}_{\alpha g}(\boldsymbol{x}^{(k)} - \alpha \nabla f(\boldsymbol{x}^{(k)}))$$

  - $\mathbf{prox}_{\alpha g}$ is easily computed as

$$\mathbf{prox}_{\alpha g}(\boldsymbol{s}) = (\boldsymbol{s} - \alpha \mathbf{1})_+ - (-\boldsymbol{s} - \alpha \mathbf{1})_+$$

- FISTA (fast ISTA) = accelerated ISTA [Beck & Teboulle '09]

group LASSO

- LASSO: [Tibshirani '96]

$$\text{Minimize} \quad \frac{1}{2}\|A\boldsymbol{x} - \boldsymbol{b}\|_2^2 + \gamma \sum_{j=1}^{n} |x_j|$$

  - attempts to zero many entries x_j
  - $\sum_{j=1}^{n} |x_j|$: ℓ_1-norm of $(x_1, \dots, x_n)$

- group LASSO (≃ von Mises): [Yuan & Lin '06]

$$\text{Minimize} \quad \frac{1}{2}\|A\boldsymbol{x} - \boldsymbol{b}\|_2^2 + \gamma \sum_{l=1}^{m} \|\boldsymbol{x}_l\|_2$$

  - attempts to zero many sub-vectors $\boldsymbol{x}_l$
  - $\sum_{l=1}^{m} \|\boldsymbol{x}_l\|_2$: ℓ_1-norm of $(\|\boldsymbol{x}_1\|_2, \dots, \|\boldsymbol{x}_m\|_2)$
  - proximal grad. meth. for group LASSO

[Mosci, Rosasco, & Santoro '10]

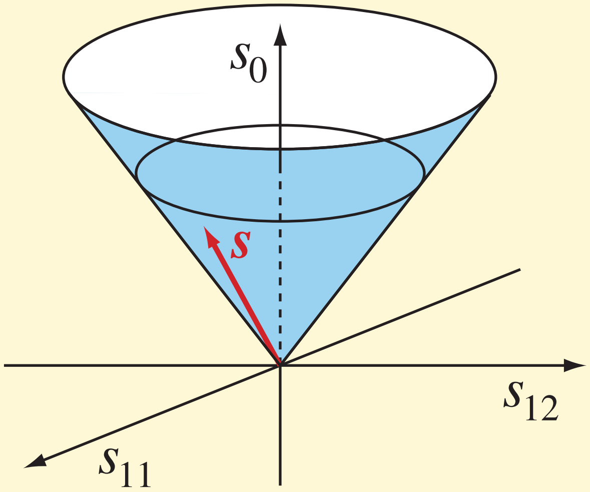
equilibrium analysis w/ von Mises criterion

- incremental problem
- optimization-based approach — potential energy minimization
  - SOCP (second-order cone programming):

    [Bisbos, Makrodimopoulos, & Pardalos '05], [Yonekura & K. '12]

Minimize	(convex quad. fcn.)
subject to	(second-order cones) & (linear eqs.)

- can be solved with a primal-dual interior-point method



$$s_0 \geq \|(s_{11}, s_{12})\|_2$$

equilibrium analysis w/ von Mises criterion

- incremental problem
- optimization-based approach — potential energy minimization
  - SOCP (second-order cone programming):
    [Bisbos, Makrodimopoulos, & Pardalos '05], [Yonekura & K. '12]

Minimize	(convex quad. fcn.)
subject to	(second-order cones) & (linear eqs.)

    - can be solved with a primal-dual interior-point method
  - equivalent formulation:

    Minimize (nonsmooth convex fcn.)

    - unconstrained optimization
    - suitable for an accelerated proximal gradient method
      - (potentially) efficient for large-scale problems

basic assumptions

- small deformation
  - strain decomposition $\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}^{\mathrm{e}} + \boldsymbol{\varepsilon}^{\mathrm{p}}$
- strain hardening
  - linear isotropic
  - linear kinematic
- incremental problem

a formulation of incremental problem

- convex, nonsmooth, unconstrained optimization:

$$\text{Min.} \quad \underbrace{\sum_{i=1}^{m} \frac{1}{2} \boldsymbol{\varepsilon}_i^{\mathrm{e}} : \mathbf{C}_i : \boldsymbol{\varepsilon}_i^{\mathrm{e}}}_{\text{elastic energy}} + \underbrace{\sum_{i=1}^{m} \left(\sqrt{\tfrac{2}{3}} R_i \|\boldsymbol{\varepsilon}_i^{\mathrm{p}}\|_{\mathrm{F}} + \frac{1}{3} \|\boldsymbol{\varepsilon}_i^{\mathrm{p}}\|_{\mathrm{F}}^2 \right)}_{\text{plastic dissipation}} - \boldsymbol{f}^{\top} \boldsymbol{u} \qquad (\spadesuit)$$

  - variables: $\boldsymbol{u}$ (inc. disp.), $\boldsymbol{\varepsilon}_i^{\mathrm{p}}$ (inc. plastic strain)
  - $\boldsymbol{\varepsilon}_i^{\mathrm{e}}$ (inc. elastic strain) can be eliminated.
    - substitute $\boldsymbol{\varepsilon}_i^{\mathrm{e}} = B_i \cdot u - \boldsymbol{\varepsilon}_i^{\mathrm{p}}$

a formulation of incremental problem

- convex, nonsmooth, unconstrained optimization:

$$\text{Min.} \quad \underbrace{\sum_{i=1}^{m} \frac{1}{2} \boldsymbol{\varepsilon}_i^{\mathrm{e}} : \mathbf{C}_i : \boldsymbol{\varepsilon}_i^{\mathrm{e}}}_{\text{elastic energy}} + \underbrace{\sum_{i=1}^{m} \left(\sqrt{\tfrac{2}{3}} R_i \|\boldsymbol{\varepsilon}_i^{\mathrm{p}}\|_{\mathrm{F}} + \frac{1}{3} \|\boldsymbol{\varepsilon}_i^{\mathrm{p}}\|_{\mathrm{F}}^2 \right)}_{\text{plastic dissipation}} - \boldsymbol{f}^{\top} \boldsymbol{u} \qquad (\spadesuit)$$

  - variables: $\boldsymbol{u}$ (inc. disp.), $\boldsymbol{\varepsilon}_i^{\mathrm{p}}$ (inc. plastic strain) $\quad \boldsymbol{\varepsilon}_i^{\mathrm{e}} := \boldsymbol{\varepsilon}_i^{\mathrm{e}}(\boldsymbol{u}, \boldsymbol{\varepsilon}_i^{\mathrm{p}})$

- apply proximal gradient method to:

$$(\spadesuit) \quad \Leftrightarrow \quad \text{Min. } f(\boldsymbol{u}, \boldsymbol{\varepsilon}^{\mathrm{p}}) + g(\boldsymbol{\varepsilon}^{\mathrm{p}}) \qquad \text{w/ } g(\boldsymbol{\varepsilon}^{\mathrm{p}}) = \sum_{i=1}^{m} \sqrt{\tfrac{2}{3}} R_i \|\boldsymbol{\varepsilon}_i^{\mathrm{p}}\|_{\mathrm{F}}$$

  - f : convex quadratic $\quad g$: convex, nonsmooth, & simple
  - $\simeq$ group LASSO:

$$\text{Min.} \quad \frac{1}{2} \|A\boldsymbol{x} - \boldsymbol{b}\|_2^2 + \gamma \sum_{l=1}^{m} \|\boldsymbol{x}_l\|_2$$

sketch of algorithm

- iteration of prox. grad. meth.:

$$\begin{bmatrix} \boldsymbol{u}^{(k+1)} \\ \boldsymbol{p} \end{bmatrix} := \begin{bmatrix} \boldsymbol{u}^{(k)} \\ \boldsymbol{\varepsilon}^{\mathrm{p}(k)} \end{bmatrix} - \alpha \nabla^2 f(\boldsymbol{u}, \boldsymbol{\varepsilon}) \begin{bmatrix} \boldsymbol{u}^{(k)} \\ \boldsymbol{\varepsilon}^{\mathrm{p}(k)} \end{bmatrix}$$

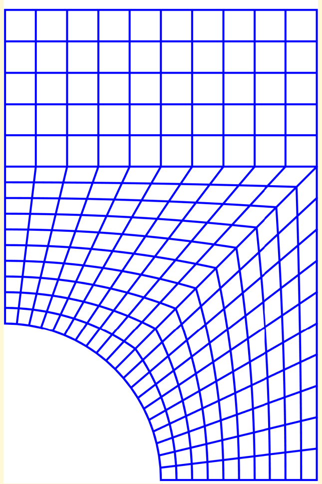
$$\boldsymbol{\varepsilon}_i^{\mathrm{p}(k+1)} := \begin{cases} \boldsymbol{0} & \text{if } \|\boldsymbol{p}_i\| \le \alpha \sqrt{\frac{2}{3}} R_i \\ \left(\|\boldsymbol{p}_i\| - \alpha \sqrt{\frac{2}{3}} R_i \right) \dfrac{\boldsymbol{p}_i}{\|\boldsymbol{p}_i\|} & \text{otherwise} \end{cases} \qquad (\forall i)$$

  - $\nabla^2 f(\boldsymbol{u}, \boldsymbol{\varepsilon})$: a sparse constant matrix
  - $\boldsymbol{p}$: intermediate variable (for notational simplicity)

- very cheap computation per iteration
- only small modification to achieve acceleration

preliminary numerical experiments

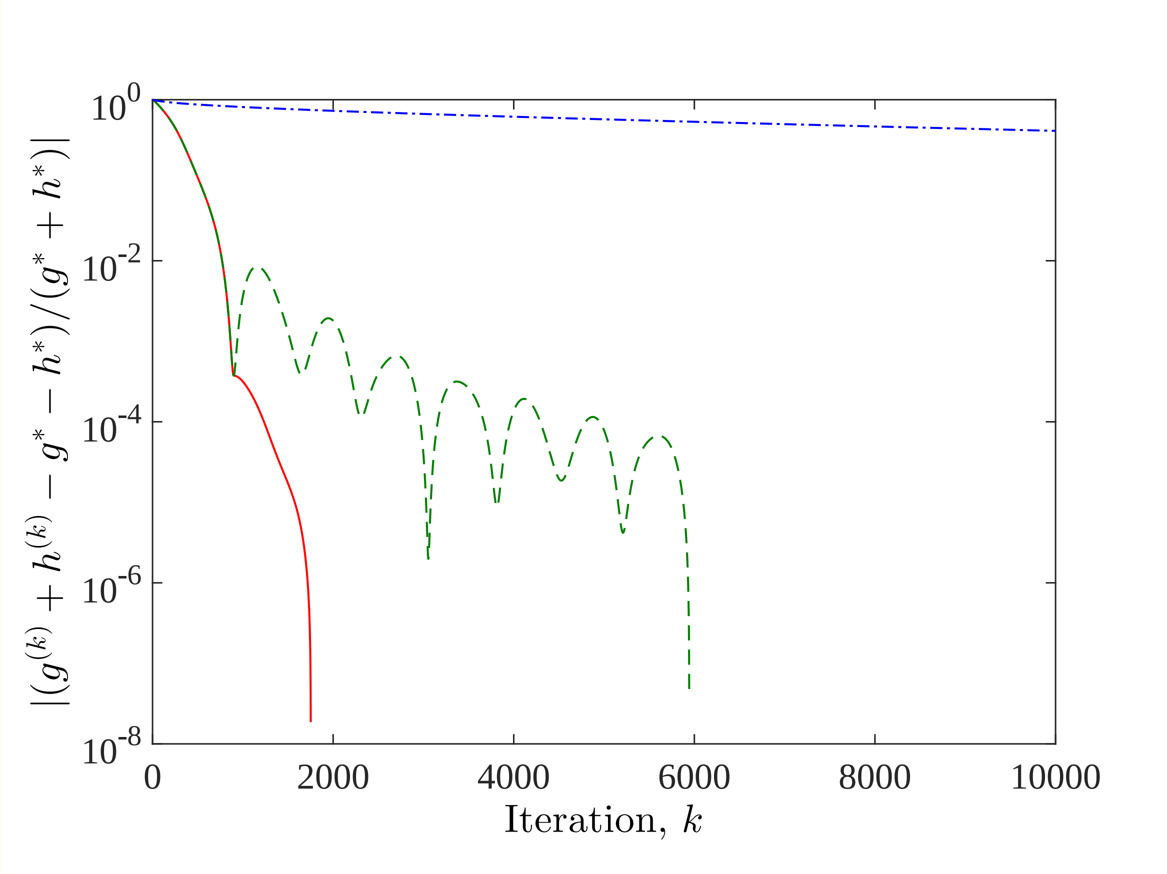
- perforated plate in 3D



- comparison
  - primal-dual interior-point method [Tütüncü, Toh, & Todd '09]
    for second-order cone programming (SOCP): SDPT3 (ver. 4)
    - solves standard form of SOCP.
  - proposed method
    - solves unconstrained nonsmooth convex opt.
    - Matlab implementation

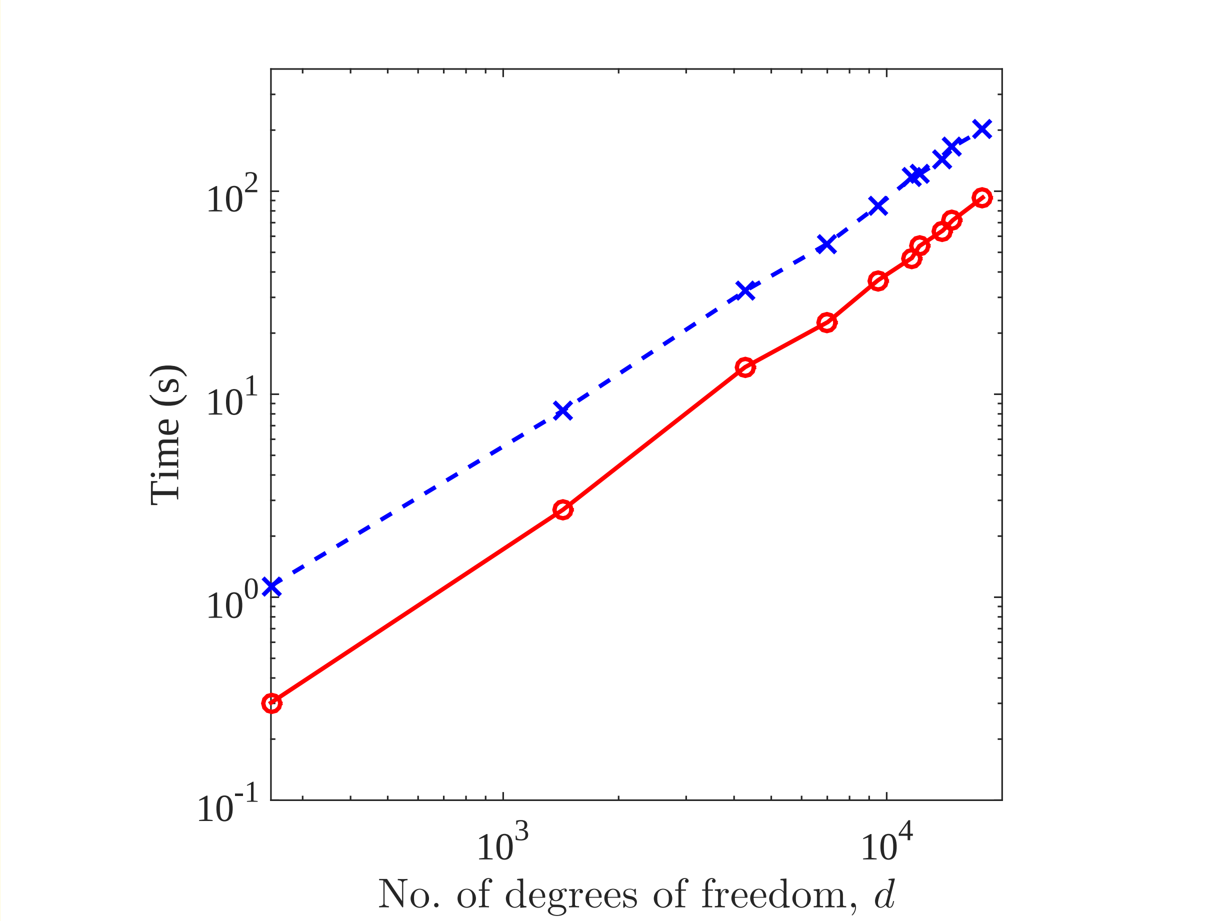
ex.) convergence history

- proximal gradient methods



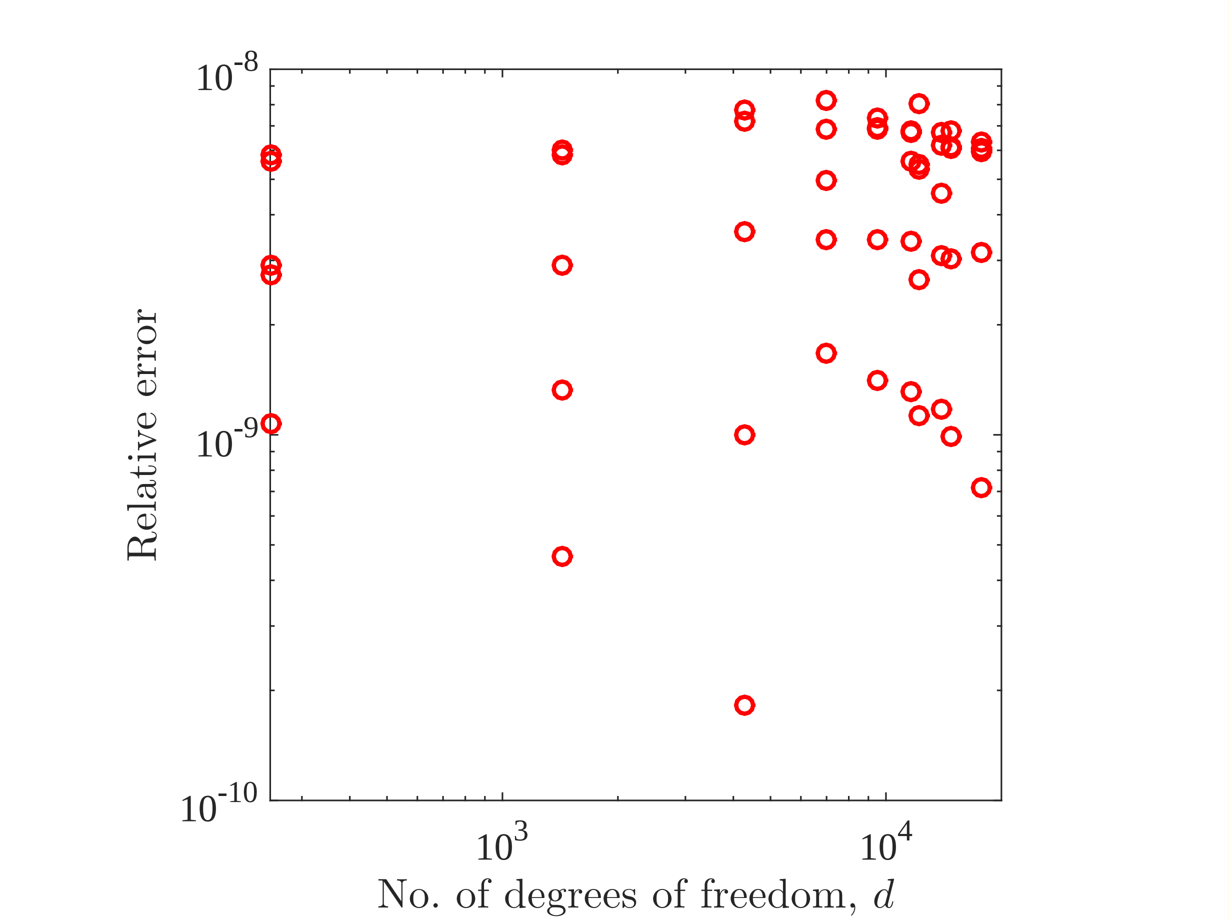
- "- - -" unaccelerated
- "……" accelerated, w/o restart
- "——" accelerated, w/ restart

ex.) computational time



- time vs. DOF
- "×" SDPT3 (interior-point method)
- "○" proposed method

ex.) computational accuracy



- objective value
  - (APGM: proposed) < (SDPT3)

- relative difference = $\dfrac{\text{(SDPT3)} - \text{(APGM)}}{\text{(SDPT3)}}$

conclusions

- incremental elastoplastic analysis w/ von Mises criterion
  - 2nd-order cone prog. (SOCP) $+$ intr.-point meth. (IPM) (exist)
  - unconstrained nonsmooth convex optimization (new)
    - (convex quadratic fcn.) $+$ (sum of ℓ_2-norms)
    - $\simeq$ group LASSO (a regularized least squares)

- accelerated proximal gradient method
  - fast convergence: $O(1/k^2)$
  - cheap computational cost (no system of linear eqs.)
    - fut. wrk.: interpret update & acceleration schemes
  - faster than a standard IPM
    - fut. wrk.: comparison with, e.g., a return-mapping meth.