Second-Order Type Optimization Algorithms For Machine Learning

Zaiwen Wen

Beijing International Center For Mathematical Research
Peking University

References/Coauthors in our group or alumnus

(a) Andre

(b) 李勇锋

(c) 柳伊扬

(d) 陈子昂

(e) 赵明明

(f) 杨明瀚

- Li Yongfeng, Wen Zaiwen, Yang Chao, Yuan Yaxiang; **A Semi-smooth Newton Method For semidefinite programs and its applications in electronic structure calculations**; SIAM Journal on Scientific Computing
- Chen Ziang, Andre Milzarek, Wen Zaiwen; **A Trust-Region Method For Nonsmooth Nonconvex Optimization**, arXiv: 2002.08513
- Andre Milzarek, Xiao Xiantao, Cen Sicong, Wen Zaiwen, Michael Ulbrich; **A stochastic semi-smooth Newton method for nonsmooth nonconvex optimization**, SIAM Journal on Optimization
- Yang Minghan, Andre Milzarek, Wen Zaiwen, Zhang Tong, **Stochastic semi-smooth Quasi-Newton method for nonsmooth optimization**
- Zhao Mingming, Li Yongfeng, Wen Zaiwen, **A stochastic trust region framework for policy optimization**

Outline

1. Basic Concepts of Semi-smooth Newton method
2. A Trust Region Method For Nonsmooth Convex Programs
3. Stochastic Semi-smooth Newton Methods
4. A stochastic trust region method for deep reinforcement learning

Composite convex program

Consider the following composite convex program

$$\min_{x \in \mathbb{R}^n} \quad f(x) + \varphi(x),$$

where f and h are convex, f is differentiable but h may not

Many applications:

- Sparse and low rank optimization: $h(x) = \|x\|_1$ or $\|X\|_*$ and many other forms.
- Regularized risk minimization: $f(x) = \sum_i f_i(x)$ is a loss function of some misfit and φ is a regularization term.
- Constrained program: φ is an indicator function of a convex set.

A General Recipe

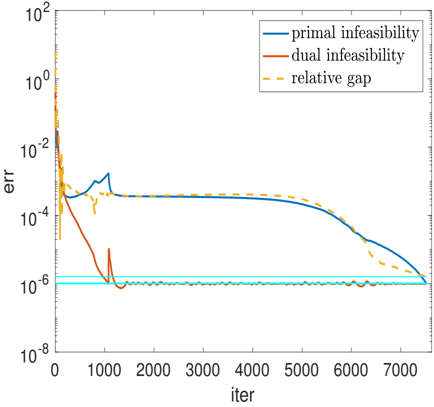
Goal: study approaches to bridge the gap between first-order and second-order type methods for composite convex programs.

key observations:

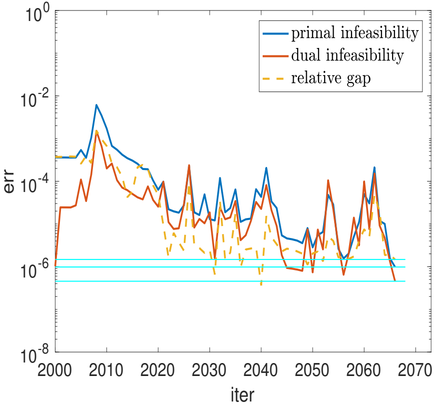
- Many popular first-order methods can be equivalent to some fixed-point iterations: $x^{k+1} = T(x^k)$;
  - **Advantages:** easy to implement; converge fast to a solution with moderate accuracy.
  - **Disadvantages:** slow tail convergence.
- The original problem is equivalent to the system $F(x) := (I - T)(x) = 0$.
- Newton-type method since $F(x)$ is semi-smooth in many cases
- Computational costs can be controlled reasonably well

An SDP From Electronic Structure Calculation

system: BeO



(g) ADMM



(h) Semi-smooth Newton

Proximal gradient method

- A first-order method

$$
\begin{aligned}
x^{k+1} &= \arg\min_x \left\langle \nabla f(x^k), x - x^k \right\rangle + \frac{\lambda}{2}\|x - x^k\|_2^2 + \varphi(x) \\
&= \operatorname{prox}_\varphi^\lambda \left(x^k - \nabla f(x^k)/\lambda \right), k = 0, 1, \cdots ,
\end{aligned}
$$

where the *proximal mapping* is:

$$
\operatorname{prox}_\varphi^\lambda(x) := \operatorname*{argmin}_{u \in \mathbb{R}^n} \{\varphi(u) + \frac{\lambda}{2}\|u - x\|_2^2\}.
$$

- Equivalent to find a root of a fixed-point mapping

$$
x = T(x) = \operatorname{prox}_\varphi^\lambda(x - \nabla f(x)/\lambda)
$$

Semi-smoothness

- Solving the system
$$F(z) = 0,$$
where $F(z) = T(z) - z$ and $T(z)$ is a fixed-point mapping.

- $F(z)$ fails to be differentiable in many interesting applications.

- but $F(z)$ is (strongly) semi-smooth and monotone.
  (a) F is directionally differentiable at x; and

  (b) for any $d \in \mathbb{R}^n$ and $J \in \partial F(x+d)$,
  $$\|F(x+d) - F(x) - Jd\|_2 = o(\|d\|_2) \quad \text{as } d \to 0.$$

A regularized semi-smooth Newton method

- The Jacobian $J_k \in \partial_B F(z^k)$ is positive semidefinite
- Let $\mu_k = \lambda_k \|F^k\|_2$. Constructe a Newton system:

$$(J_k + \mu_k I)d = -F^k,$$

- Solving the Newton system inexactly:

$$r^k := (J_k + \mu_k I)d^k + F^k.$$

  We seek a step d^k approximately such that

$$\|r^k\|_2 \le \tau \min\{1, \lambda_k \|F^k\|_2 \|d^k\|_2\}, \quad \text{where } 0 < \tau < 1$$

- Newton Step: $z^{k+1} = z^k + d^k$
- Faster local convergence is ensured

Semidefinite Programming

Consider the SDP

$$\min \langle C, X \rangle, \text{s.t.} \quad \mathcal{A}X = b, X \succeq 0$$

- $f(X) = \langle C, X \rangle + 1_{\{\mathcal{A}X=b\}}(X)$.
- $h(X) = 1_K(X)$, where $K = \{X : X \succeq 0\}$.
- Proximal Operator: $\mathrm{prox}_{th}(Z) = \arg\min_X \frac{1}{2}\|X - Z\|_F^2 + th(X)$
- Let $Z = Q\Sigma Q^T$ be the spectral decomposition

$$\begin{aligned}\mathrm{prox}_{tf}(Y) &= (Y + tC) - \mathcal{A}^*(\mathcal{A}Y + t\mathcal{A}C - b), \\ \mathrm{prox}_{th}(Z) &= Q_\alpha \Sigma_\alpha Q_\alpha^T,\end{aligned}$$

- Fixed-point mapping from DRS:

$$F(Z) = \mathrm{prox}_{th}(Z) - \mathrm{prox}_{tf}(2\mathrm{prox}_{th}(Z) - Z) = 0.$$

Semi-smooth Newton System

- assumption: $\mathcal{A}\mathcal{A}^* = I$
- The SMW theorem yields the inverse matrix

$$(J_k + \mu_k I)^{-1} = H^{-1} + H^{-1}A^T(I - AWH^{-1}A^T)^{-1}AWH^{-1}$$

$$= \frac{1}{\mu(\mu+1)}(\mu I + T)(I + A^\top(\frac{\mu^2}{2\mu+1}I + ATA^\top)^{-1}A(\frac{\mu}{2\mu+1}I - T)).$$

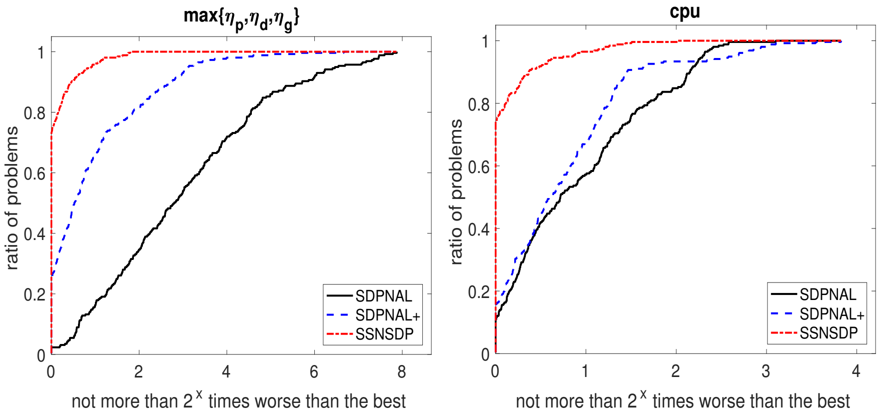
- $ATA^\top d = \mathcal{A}Q(\Omega_0 \circ (Q^T(D)Q))Q^T$, where $D = \mathcal{A}^*d$,

$$\Omega_0 = \begin{bmatrix} E_{\alpha\alpha} & l_{\alpha\bar{\alpha}} \\ l_{\alpha\bar{\alpha}}^T & 0 \end{bmatrix},$$

  and $E_{\alpha\alpha}$ is a matrix of ones and $l_{ij} = \frac{\mu k_{ij}}{\mu+1-k_{ij}}$

- computational cost $O(|\alpha|n^2)$

Comparison on electronic structure calculation



(i) max(η_p,η_d,η_g) (j) cpu time

Optimal Transport

Linear programming:

$$\begin{aligned} \min_{X \in \mathbb{R}^{m\times n}} \quad & \langle C, X\rangle, \\ \text{s.t.} \quad & \sum_{j=1}^{n} X_{i,j} = u_i, \quad 1 \le i \le m, \\ & \sum_{i=1}^{m} X_{i,j} = v_j, \quad 1 \le j \le n, \\ & X_{i,j} \ge 0, \quad 1 \le i \le m, 1 \le j \le n, \end{aligned}$$

where $C \in \mathbb{R}^{m\times n}$ is the given cost matrix.

- Sparsity
- Multilevel scheme

Squared ℓ_2-DOTmark 128×128 images

	MSSN		CPLX-NWS	M-CPLX
Class	TIME/SSN/CG	gap/pinf/dinf	TIME	TIME
WhiteNoise	24.86/1717/18839	3.57e-07/9.90e-07/2.98e-08	1262.96	22.09
GRFrough	21.61/1375/12727	2.00e-07/7.28e-07/4.20e-08	1398.86	53.71
GRFmod	18.28/1049/8573	1.14e-09/9.69e-07/1.19e-07	1703.69	51.16
GRFsmooth	35.15/1467/17149	1.79e-08/9.86e-07/3.45e-08	1892.41	69.25
LogGRF	94.41/3945/22768	2.23e-10/9.93e-07/7.83e-07	2066.44	56.17
LogitGRF	83.57/3276/33599	1.31e-08/8.96e-07/9.57e-07	1928.92	83.84
Cauchy	104.64/17826/256255	1.86e-07/9.65e-07/9.34e-07	1869.37	51.30
Shapes	9.12/748/3380	1.19e-08/5.67e-07/3.38e-10	2501.76	12.11
Classic	31.73/2820/27321	1.18e-07/7.45e-07/3.27e-07	1732.93	70.36
Microscopy	24.69/1663/10880	8.52e-09/9.98e-07/9.30e-08	1671.90	35.14

Outline

Problem setup

- Nonsmooth composite program:

$$\min_{x\in\mathbb{R}^n} \psi(x) := f(x) + \varphi(x),$$

  where $f:\mathbb{R}^n \to \mathbb{R}$ is a (probably nonconvex) smooth function and $\varphi:\mathbb{R}^n \to \mathbb{R}$ is a convex, proper, and lower semi-continuous mapping.
- Trust-region subproblem:

$$\min_{s\in\mathbb{R}^n} m_k(p) = \psi_k + g_k^T p + \frac{1}{2}p^T B_k p, \quad \text{s.t.} \quad \|p\| \le \Delta_k.$$

- $g(x)$ is an extension of the gradient and will be constructed later.
- A desired property: $m_k(p)$ locally fits $\psi(x)$ well along a specific direction.

Construction of $g(x)$

- The steepest descent direction: $d_s(x) = \underset{d \in \mathbb{R}^n,\ \|d\| \leq 1}{\operatorname{argmin}} \ \psi'(x; d)$.
- In the smooth case: $\nabla \psi(x) = \psi'(x; d_s(x)) d_s(x)$.
- In the nonsmooth case, we choose a descent direction $d(x)$ with

$$\|d(x)\| = \begin{cases} 0, & 0 \in \partial \psi(x), \\ 1, & 0 \notin \partial \psi(x), \end{cases}$$

  and an upper bound of the directional derivative:

$$u(x) \in \begin{cases} [\psi^o(x, d(x)), 0), & 0 \notin \partial \psi(x), \\ \{0\}, & 0 \in \partial \psi(x). \end{cases}$$

- $g(x) := u(x) d(x)$.

Preferable Choices of $d(x)$ and $u(x)$

Choice 1:

- We say $d_\gamma(x)$ is a γ-*inexact steepest descent direction* ($\gamma \in (0,1]$) if it satisfies $\|d_\gamma(x)\| \leq 1$ and $\psi'(x; d_\gamma(x)) \leq \gamma \psi'(x; d_s(x))$.
- $d(x) = d_\gamma(x)$, $u(x) = \psi'(x; d_\gamma(x))$.
- Choice 1 may be difficult to implement.

Choice 2:

- *Proximal Operator*: $\text{prox}_\varphi^\Lambda(x) := \operatorname*{argmin}_{z \in \mathbb{R}^n} \varphi(z) + \frac{1}{2}\|z - x\|_\Lambda^2$.
- *Natural Residual*: $F_{\text{nat}}^\Lambda(x) := x - \text{prox}_\varphi^\Lambda(x - \Lambda^{-1}\nabla f(x))$.
- A point x^* is a stationary point of problem (16) if and only if x^* is a solution of the nonsmooth equation $F_{\text{nat}}^\Lambda(x) = 0$.
- $\psi'\left(x; -F_{\text{nat}}^\Lambda(x)\right) \leq -\left\|F_{\text{nat}}^\Lambda(x)\right\|_\Lambda^2$.
- $d(x) = -\frac{F_{\text{nat}}^\Lambda(x)}{\left\|F_{\text{nat}}^\Lambda(x)\right\|}$, $u(x) = -\lambda_{\min}\left\|F_{\text{nat}}^\Lambda(x)\right\|$.

Model Function and Trust-Region Subproblem

- Let $g_k = u(x_k)d(x_k)$. Trust region subproblem:

$$\min_s \ m_k(s) = \psi_k + \langle g^k, s\rangle + \frac{1}{2}\langle s, B^k s\rangle \quad \text{s.t.} \quad \|s\| \le \Delta_k$$

- Cauchy point: $p_k^C := -\alpha_k^C g_k$ and $\alpha_k^C := \operatorname*{argmin}_{0\le t\le \frac{\Delta_k}{\|g_k\|}} m_k(-tg_k)$.

- Choose the regularization parameter:

$$\frac{1}{2}h^T B^k h + t_k\|h\|^2 \ge \lambda_1\|h\|^2 \quad \forall\, h\in\mathbb{R}^n \quad \text{and} \quad \|B^k + t_k I\| \le \lambda_2,$$

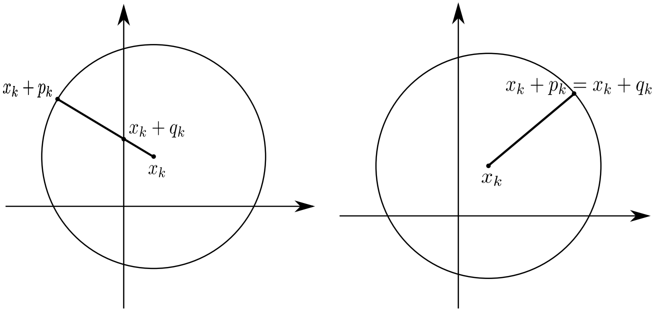
Solve a system: $(B^k + t_k I)p = -g^k$ such that

$$(B^k + t_k I)p^k = -g^k + r^k \quad \text{and} \quad \|r^k\| \le \frac{\lambda_1}{2(\lambda_1+\lambda_2)}\|g^k\|.$$

Project p^k onto the trust region: $s^k = \min\{\Delta_k, \|p^k\|\}\bar{p}^k$

Suitable Stepsize

- Descent direction $\bar{p}_k = \frac{p_k}{\|p_k\|}$.
- $\Gamma_{\max}(x, d) := \sup \left\{ T > 0 : \tilde{\psi}^o_{x,d}(t) := \psi^o(x + td; d) \in C(0, T) \right\}$
- $\Gamma(x) := \inf_{d \in \mathbb{R}^n,\ \|d\|=1} \Gamma_{\max}(x, d)$
- Stepsize $\alpha_k = \min \left\{ \Gamma\left(x_k; \bar{p}_k\right), \|p_k\| \right\}$.
- Example: $n = 2$, $\varphi(x) = \|x\|_1$, where $q_k := \alpha_k \bar{p}_k$.



Truncation Step

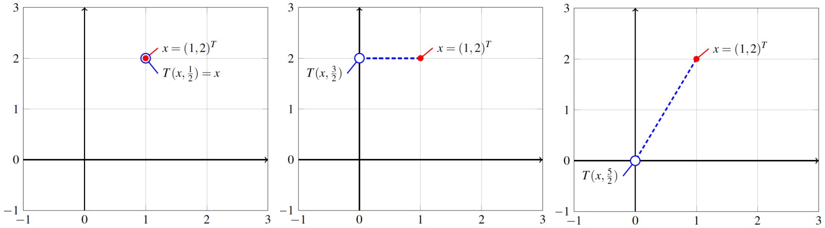
Definition 1

If there exists a sequence $\{S_i\}_{i=0}^m$ *satisfying* $\mathbb{R}^n = S_0 \supset S_1 \cdots \supset S_m$*,* $\delta \in (0, +\infty]$*,* $\kappa > 0$*, and a function* $T : \mathbb{R}^n \times (0, \delta] \to \mathbb{R}^n$ *with following properties:*

(1) $\Gamma(x) \geq \delta, \ \forall x \in S_m$*;*

(2) For any $a \in (0, \delta]$ *and* $x \in S_i \backslash S_{i+1}$ *(*$i \in 0, 1, \cdots, m-1$*), if* $\Gamma(x) \geq a$*, it holds* $T(x, a) = x$*; if* $\Gamma(x) < a$*, it holds* $T(x, a) \in S_{i+1}$*,* $\Gamma(T(x, a)) \geq a$*, and* $\|T(x, a) - x\| \leq \kappa a$*;*

we say φ *is truncatable and* T *is a truncation operator.*



Global Convergence

Assumption 1

We assume that ψ and f have the following properties:
(A.1) $\nabla f(x)$ is locally Lipschitz continuous on $\mathbb{R}^n$.
(A.2) ψ is bounded from below by L_b.

Assumption 1

Let $\{x_k\}$ and $\{B_k\}$ be generated by the Algorithm, we assume:
(B.1) $\{x_k\}_{k\in\mathbb{N}}$ is bounded, i.e., there exist $R > 0$ with $\{x_k\} \subseteq B_R(0)$.
(B.2) There exists $\kappa_B > 0$ with $\sup_{k\in\mathbb{N}} \|B_k\| \leq \kappa_B < \infty$.
(B.3) For any subsequence $\{k_\ell\}_{\ell=0}^{\infty} \subseteq \mathbb{N}$, if $\{x_{k_\ell}\}$ is convergent and $\alpha_{k_\ell} \to 0$, then we have

$$\varphi(x_{k_\ell} + \alpha_{k_\ell}\bar{s}_{k_\ell}) - \varphi(x_{k_\ell}) - \alpha_{k_\ell}\varphi^o(x^{k_\ell}; \bar{s}_{k_\ell}) \leq o(\alpha_{k_\ell}).$$

(B.4) For every $\epsilon > 0$ there is $\epsilon' > 0$ such that for all x^k with $\Gamma(x^k) \geq \epsilon$ it follows $\Gamma\left(x^k, \bar{s}^k\right) \geq \epsilon'$.

Global Convergence

Theorem 1

For truncatable φ, suppose that (A.1), (A.2), (B.1)-(B.4) are satisfied. Assume that the Algorithm does not terminate in finitely many steps and let $\{x_k\}_{k=0}^{\infty}$ be the sequence generated by the Algorithm. Then it holds that

$$\liminf_{k\to\infty} \|g_k\| = 0.$$

Theorem 1

*Under the same assumptions as in the last Theorem, let x^* be any accumulation point of the sequence $\{x_k\}_{k=0}^{\infty}$ generated by the Algorithm where g_k is given by **Choice 1** or **Choice 2**. Then x^* is an stationary point of* (16).

Outline

1. Basic Concepts of Semi-smooth Newton method
2. A Trust Region Method For Nonsmooth Convex Programs
3. Stochastic Semi-smooth Newton Methods
4. A stochastic trust region method for deep reinforcement learning

Stochatic optimization problem

- Consider

$$\min_{x\in\mathbb{R}^n} \Psi(x) := f(x) + \varphi(x)$$

- Expected and Empirical Risk Minimization:

$$f(x) := \mathbb{E}[F(x,\xi)], \qquad f(x) = \frac{1}{N}\sum_{i=1}^{N} f_i(x)$$

- Assume $f(x)$ is smooth but $\varphi(x)$ is convex and non-smooth.
- Large-scale machine learning problems: the number of data samples N is very large
- Full evaluation of $f(x)$ and $\nabla f(x)$ is not tractable or simply too expansive.

Algorithmic Idea

Basic idea based on $x^{k+1} = \mathrm{prox}_{\varphi}^{\lambda}(x^k - t\nabla f(x^k))$.

- We incorporate second order information and use stochastic Hessian oracles ($\mathcal{SSO}$)

$$H_{t^k}(x^k) \approx \nabla^2 f(x^k)$$

  to estimate the Hessian $\nabla^2 f$ and compute the Newton step.
- The sample collections s^k and t^k are chosen independently of each other and of the other batches s^ℓ, t^ℓ, $\ell \in \mathbb{N}_0 \setminus \{k\}$.
- We work with the following $\mathcal{SFO}$ and $\mathcal{SSO}$:

$$\nabla f_{s^k}(x) := \frac{1}{|s^k|} \sum_{i \in s^k} \nabla f_i(x) \quad \text{and} \quad \mathcal{H}_{t^k}(x) := \frac{1}{|t^k|} \sum_{i \in t^k} \nabla^2 f_i(x).$$

Stochastic Semi-smooth Newton Method: Idea

To accelerate the stochastic proximal gradient method, we want to augment it by a stochastic Newton-type step, obtained from the (sub-sampled) optimality condition:

$$F_s^\lambda(x) = x - \mathrm{prox}_h^\lambda(x - \lambda^{-1}\nabla f_s(x)) \approx 0.$$

The semi-smooth Newton step is given by

$$M_k\, d^k = -F_{s^k}^\lambda(x^k), \quad x^{k+1} = x^k + d^k,$$

with sample batches s^k, t^k and $M_k \in \mathcal{M}_{s^k,t^k}^{\lambda_k}(x^k)$,

$$\mathcal{M}_{s,t}^\lambda(x) := \{M = I - D + D\lambda^{-1}\mathcal{H}_t(x) : D \in \partial\mathrm{prox}_\varphi^\lambda(u_s^\lambda(x))\}$$

and $u_s^\lambda(x) := x - \lambda^{-1}\nabla f_s(x)$.

$\rightsquigarrow$ Aim: Utilize fast local convergence to stationary points!

Algorithmic Framework

We use the following growth conditions $(\star)$:

$$\|F^{\lambda_{k+1}}_{s^{k+1}}(z^k)\| \leq (\eta + \nu_k) \cdot \theta_k + \varepsilon_k^1, \tag{G.1}$$

$$\psi(z^k) \leq \psi(x^k) + \beta \cdot \theta_k^{1/2} \|F^{\lambda_{k+1}}_{s^{k+1}}(z^k)\|^{1/2} + \varepsilon_k^2, \tag{G.2}$$

where $\eta \in (0,1)$, $\beta > 0$, and $(\nu_k), (\varepsilon_k^2) \in \ell_+^1$, $(\varepsilon_k^1) \in \ell_+^{1/2}$.

We set θ_{k+1} to $\|F^{\lambda_{k+1}}_{s^{k+1}}(x^{k+1})\|$ if x^{k+1} was obtained in step 3.

Remark:

- Calculating $F^{\lambda_{k+1}}_{s^{k+1}}(z^k)$ requires evaluation of $\nabla f_{s^{k+1}}(z^k)$. This information can be reused in the next iteration if $z^k \rightsquigarrow x^{k+1}$ is accepted as new iterate.

Global Convergence: Assumptions

Basic Assumptions:

(A.1) ∇f is Lipschitz continuous on $\mathbb{R}^n$ with constant L.

(A.2) The matrices $(\lambda_k) \subset \mathbb{S}^n_{++}$ satisfy $\lambda_M I \succeq \lambda_k \succeq \lambda_m I$ for all k.

(A.3) ψ is bounded from below on $\mathbf{dom}\ \varphi$.

Stochastic Assumptions:

(S.1) For all $k \in \mathbb{N}$, there exists $\sigma_k \geq 0$ such that

$$\mathbb{E}[\|\nabla f(x^k) - \nabla f_{s^k}(x^k)\|^2] \leq \sigma_k^2.$$

(S.2) The matrices M_k, chosen in step 1, are random operators.

Global Convergence

Theorem: Global Convergence [*MXCW*, '17]

Suppose that (A.1)–(A.3) and (S.1)–(S.2) are fulfilled. Then, under the additional conditions, $\alpha_k \leq \overline{\alpha} := \min\{1, \lambda_m/L\}$,

$$(\alpha_k) \text{ is nonincreasing }, \quad \sum \alpha_k = \infty, \quad \sum \alpha_k \sigma_k^2 < \infty$$

it holds $\liminf_{k\to\infty} \mathbb{E}[\|F^\lambda(x^k)\|^2] = 0$ and $\liminf_{k\to\infty} F^\lambda(x^k) = 0$ a.s. for any $\lambda \in \mathbb{S}^n_{++}$.

- Verify that (x^k) actually defines an adapted stochastic process.
- The batch s^k and the iterate x^k are not independent.
- Derive approximate and uniform descent estimates for the terms $\psi(x^k) - \psi(x^{k+1})$.

For strongly convex case: $\lim_{k\to\infty} \mathbb{E}[\|F^\lambda(x^k)\|^2] = 0$ and $\lim_{k\to\infty} F^\lambda(x^k) = 0$ a.s. for any $\lambda \in \mathbb{S}^n_{++}$.

Stochastic Semi-smooth Quasi-Newton Method

- Use stochastic approximation technique!
  Estimate $v^k \approx \nabla f(x^k)$ from stochastic oracle and set

$$F_{v^k}(x^k) := x^k - \mathrm{prox}_{\varphi}^{\lambda}(x^k - v^k/\lambda).$$

  Example: Assume the samples s are chosen independently, then a possible estimate of $\nabla f(x)$ is $\nabla f_s(x^k) := \frac{1}{|s|}\sum_{i\in s} \nabla f_i(x^k)$.
- Use extra-gradient step for globalization!
  (a) First employ the "Newton" step:

$$z^k = x^k + \beta_k d^k, \quad d^k = -W^k F_{v^k}(x^k)$$

  where W^k is exact or approximation of inverse of J^k.

  (b) Perform an extra gradient step:

$$x^{k+1} = \mathrm{prox}_{\varphi}^{\lambda}(x^k + \alpha_k d^k - v_+^k/\lambda), \quad v_+^k \approx \nabla f(z^k).$$

The choice of β_k and α_k are very flexible !

Coordinate Quasi-Newton Method

- Further computation reduction?
  Use coordinate update!
- Given a coordinates set $\mathcal{A}(x^k)$ and $\mathcal{O}(x^k) := [N] \setminus \mathcal{A}(x^k)$, d^k is updated by coordinate set:

$$d^k = -\begin{bmatrix} W_{\mathcal{A}(x^k)\mathcal{A}(x^k)} & 0 \\ 0 & \gamma_k I \end{bmatrix} \begin{bmatrix} (F^{\lambda}_{\nu^k}(x^k))_{\mathcal{A}(x^k)} \\ (F^{\lambda}_{\nu^k}(x^k))_{\mathcal{O}(x^k)} \end{bmatrix},$$

- $W_{\mathcal{A}(x^k)\mathcal{A}(x^k)}$ is updated by L-BFGS related to coordinates $\mathcal{A}(x^k)$.

$$(U^k)_{\mathcal{A}(x^k)} = [u^{k-p}_{\mathcal{A}(x^k)}, \ldots, u^{k-1}_{\mathcal{A}(x^k)}], \quad (Y^k)_{\mathcal{A}(x^k)} = [y^{k-p}_{\mathcal{A}(x^k)}, \ldots, y^{k-1}_{\mathcal{A}(x^k)}],$$

are the subvectors of U^k, Y^k.

Convergence Assumption

Basic Assumption

- A.1 The gradient mapping ∇f is Lipschitz continuous on $\mathbb{R}^n$ with modulus $L_f \geq 1$.
- A.2 The objective function ψ is bounded from below on $\mathbf{dom}\ \varphi$.
- A.3 $\varphi : \mathbb{R}^n \to (-\infty, \infty]$ is convex, lower semicontinuous, and proper.

Stochastic Assumption

- B.1 The mapping $\mathsf{D}^k : \Omega \to \mathbb{R}^n$ is an $\mathcal{F}^k$-measurable function for all k.
- B.2 There is $\nu_k > 0$ such that we have $\mathbb{E}[\|\mathsf{D}^k\|^2 \mid \mathcal{F}_+^{k-1}] \leq \nu_k^2 \cdot \mathbb{E}[\|F_{\mathsf{V}^k}(\mathsf{X}^k)\|^2 \mid \mathcal{F}_+^{k-1}]$ a.e. and for all $k \in \mathbb{N}$.
- B.3 For all $k \in \mathbb{N}$, it holds $\mathbb{E}[\mathsf{V}^k \mid \mathcal{F}_+^{k-1}] = \nabla f(\mathsf{X}^k)$, $\mathbb{E}[\mathsf{V}_+^k \mid \mathcal{F}^k] = \nabla f(\mathsf{Z}^k)$ a.e. and there exists $\sigma_k, \sigma_{k,+} > 0$ such that a.e.

$$\mathbb{E}[\|\nabla f(\mathsf{X}^k) - \mathsf{V}^k\|^2 \mid \mathcal{F}_+^{k-1}] \leq \sigma_k^2 \quad \text{and} \quad \mathbb{E}[\|\nabla f(\mathsf{Z}^k) - \mathsf{V}_+^k\|^2 \mid \mathcal{F}^k] \leq \sigma_{k,+}^2,$$

where

$$\mathcal{F}^k = \sigma(V^0, V_+^0, \dots, V^k) \quad \text{and} \quad \mathcal{F}_+^k = \sigma(\mathcal{F}_k \cup \sigma(V_+^k)).$$

Theorem 1

Suppose that the assumptions (A.1)–(A.3) *and* (B.1)–(B.3) *are satisfied and we have*

$$\lambda_{k,+} \leq \frac{1}{L_f}, \quad \lambda_k \leq \frac{(1-\bar{\rho})\lambda_{k,+}}{1+\mu_k^2},$$

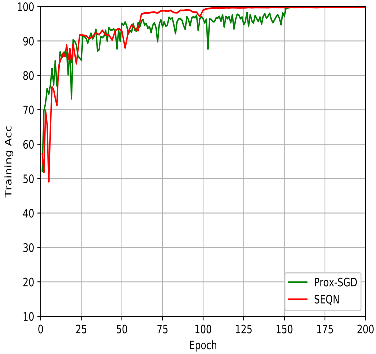
where $\mu_k = \nu_k(\alpha_k + L_f \beta_k \lambda_{k,+})$*. Then, under the additional conditions*

$$\sum \lambda_k = \infty, \quad \sum \lambda_k \sigma_k^2 < \infty, \quad \sum \lambda_{k,+} \sigma_{k,+}^2 < \infty$$

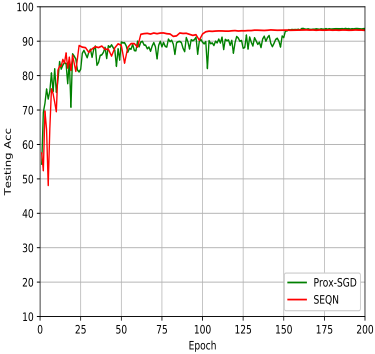
it follows $\liminf_{k\to\infty} \mathbb{E}[\|F(\mathsf{X}^k)\|^2] = 0$ *and* $\liminf_{k\to\infty} F(\mathsf{X}^k) = 0$ *a.s. and* $(\psi(\mathsf{X}^k))_k$ *a.s. converges to some random variable* Y^* *with* $\lim_{k\to\infty} \mathbb{E}[\psi(\mathsf{X}^k)] = \mathbb{E}[\mathsf{Y}^*]$*.*

Locally*, if we further assume the function satisfy KL-property and some mild assumption, we can show then* $(\mathsf{X}^k)_k$ *converges almost surely to a* $\mathrm{crit}\,\psi$*-valued random variable* X^**.*

Deep learning: ResNet-18 on Cifar10, $\psi(x) = \|x\|_1$



(a) Training accuracy

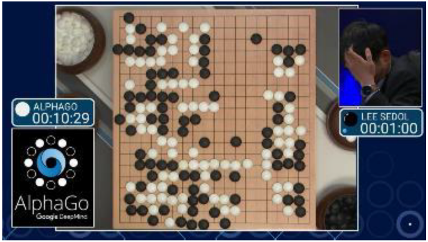


(b) Testing accuracy

Outline

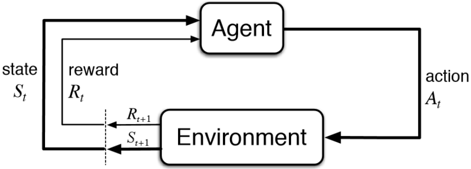
1. Basic Concepts of Semi-smooth Newton method
2. A Trust Region Method For Nonsmooth Convex Programs
3. Stochastic Semi-smooth Newton Methods
4. A stochastic trust region method for deep reinforcement learning

Reinforcement learning



Preliminaries

- Consider an infinite-horizon discounted Markov decision process (MDP), usually defined by a tuple $(\mathcal{S}, \mathcal{A}, P, R, \rho_0, \gamma)$;



  - ρ_0: the distribution of s_0
  - γ: discount factor $\in (0, 1)$
  - P: transition probability

- A trajectory: $\tau = \{s_0, a_0, r(s_0, a_0), s_1, ..., s_t, a_t, r(s_t, a_t), s_{t+1}, ...\}$.
- At a given state, choose action from $\pi(\cdot|s)$: $\int_{\mathcal{A}} \pi(a|s)da = 1$.
- The policy is supposed to maximize the total expected reward:

$$\max_{\pi} \ \eta(\pi) = \mathrm{E}_{\pi}\left[\sum_{t=0}^{\infty} \gamma^t r(s_t, a_t)\right],$$

with $s_0 \sim \rho_0, a_t \sim \pi(\cdot|s_t), s_{t+1} \sim P(\cdot|s_t, a_t)$.

Deep reinforcement learning

- In real-world tasks: high dimensionality, limited observations,...
- In deep reinforcement learning, the policy π and/or value functions are usually parameterized with differentiable neural networks.
- The policy-based optimization:

$$\max_{\theta} \quad \eta(\theta).$$

- The value-based optimization:

$$\min_{\phi} \mathrm{E}_{s,a} \left\{ Q_{\phi}(s,a) - \mathrm{E}_{s' \sim \mathrm{P}(\cdot|\mathrm{s,a})} \left[r(s,a) + \gamma \max_{a'} Q_{\phi}(s',a')|s,a \right] \right\}^2 .$$

- Challenges: theoretical analysis; generalization; stability; trade off between exploration and exploitation...

VPG, NPG

- Policy gradient: $\nabla\eta(\theta) = \mathbb{E}_{\rho_\theta,\pi_\theta}\left[\nabla \log \pi_\theta(a|s) A_\theta(s,a)\right]$.
- $\rho_\theta(s) = \sum_{t=0}^{\infty} \gamma^t P(s_t = s|\pi_\theta)$ is the (unnormalized) discounted visitation frequencies.
- Vanilla[1]/Natural[2] policy gradient: $\theta_{k+1} = \theta_k + \alpha M(\theta_k)\nabla_\theta\eta(\theta_k)$.
- $M(\theta_k)^{-1} = \mathbb{E}_{\rho_{\theta_k},\pi_{\theta_k}}\left[\nabla_\theta \log \pi_{\theta_k}(s,a) \nabla_\theta \log \pi_{\theta_k}(s,a)^T\right]$.
- A local approximation of η:

$$\eta(\theta) = \eta(\theta_k) + \sum_s \rho_\theta(s) \sum_a \pi_\theta(a|s) A_{\theta_k}(s,a),$$

$$L_{\theta_k}(\theta) = \eta(\theta_k) + \sum_s \rho_{\theta_k}(s) \sum_a \pi_\theta(a|s) A_{\theta_k}(s,a).$$

- $\eta(\theta_k) = L_{\theta_k}(\theta_k), \nabla\eta(\theta_k) = \nabla L_{\theta_k}(\theta_k)$.

[1] R. S. Sutton, el al., Policy gradient methods for reinforcement learning with function approximation.

[2] S. M. Kakade, A natural policy gradient.

Stochastic Trust Region Algorithm

- The objective function
$$\max_{\theta} \quad \eta(\theta).$$
- At k-th iteration, obtain a trail point $\tilde{\theta}_{k+1}$ from the subproblem:
$$\max_{\theta} \quad L_{\theta_k}(\theta), \quad \text{s.t. } \mathbb{E}_{s\sim\rho_{\theta_k}}\left[D(\pi_{\theta_k}(\cdot|s), \pi_{\theta}(\cdot|s))\right] \le \delta_k.$$
- Compute the ratio $r_k = \frac{\eta(\tilde{\theta}_{k+1}) - \eta(\theta_k)}{L_{\theta_k}(\tilde{\theta}_{k+1}) - L_{\theta_k}(\theta_k)}$.
- Update $\theta_{k+1} = \begin{cases} \tilde{\theta}_{k+1}, & r_k \ge \beta_0, \\ \theta_k, & \text{o.w.}, \end{cases}$, with $\beta_0 > 0$.
- Update $\delta_{k+1} = \mu_{k+1}\|\nabla L_{\theta_{k+1}}(\theta_{k+1})\|$ with $\gamma_1 > 1 \ge \gamma_2 > \gamma_3$,
$$\mu_{k+1} = \begin{cases} \gamma_1\mu_k, & r_k \ge \beta_1, \\ \gamma_2\mu_k, & r_k \in [\beta_0, \beta_1), \\ \gamma_3\mu_k, & \text{o.w.}, \end{cases}.$$

Unparameterized Policy

- Specifying the total variation distance in discrete cases (the KL divergence in continuous cases).
- Policy advantage: $\mathbb{A}_\pi(\pi') = \mathrm{E}_{s\sim\rho_\pi}\left[\mathrm{E}_{a\sim\pi'(\cdot|s)}\left[A_\pi(s,a)\right]\right]$.

Lemma 2 (Optimality condition)

π is the optimal policy if and only if

$$\mathbb{A}_\pi^* = \max_{\pi'} \mathbb{A}_\pi(\pi') = 0, \text{ i.e., } \pi \in \mathrm{argmax}_{\pi'} \mathbb{A}_\pi(\pi').$$

Lemma 3 (Monotonicity)

Suppose $\{\pi_k\}$ is the sequence generated by our trust region method, then we have $\eta(\pi_{k+1}) \geq \eta(\pi_k)$, the equality holds if and only if π_k is the optimal policy.

Main Results

Lemma 4 (Lower bound of ΔL_{π_k})

Suppose $\{\pi_k\}$ is the sequence generated by our trust region method, then we have $L_{\pi_k}(\pi_{k+1}) - L_{\pi_k}(\pi_k) \geq \min(1, (1-\gamma)\delta_k)\mathbb{A}^*_{\pi_k}$.

Lemma 5 (Lower bound of r_k)

The ratio r_k satisfies that $r_k \geq \min\left(1 - \frac{4\epsilon_k\gamma\delta_k^2}{p_0^2(1-\gamma)^2\mathbb{A}^*_{\pi_k}}, 1 - \frac{4\epsilon_k\gamma\delta_k}{p_0^2(1-\gamma)^3\mathbb{A}^*_{\pi_k}}\right)$, where $p_0 = \min_s \rho_0(s)$ and $\epsilon_k = \max_{s,a}|A_{\pi_k}(s,a)|$.

Theorem 6 (Convergence)

Suppose $\{\pi_k\}$ is the sequence generated by our trust region method, then we have the following conclusions

1. $\lim_{k\to\infty} \mathbb{A}^*_{\pi_k} = 0$.
2. $\lim_{k\to\infty} \eta(\pi_k) = \eta(\pi^*)$, *where π^* is the optimal policy.*

Empirical algorithm

- Terminate condition:

$$\frac{|\hat{L}_{\theta_k}(\theta_{k,l+1}) - \hat{L}_{\theta_k}(\theta_{k,l})|}{1 + \left|\hat{L}_{\theta_k}(\theta_{k,l})\right|} \leq \epsilon, \text{ or } \frac{|\mathrm{Ent}(\theta_{k,l+1}) - \mathrm{Ent}(\theta_k)|}{1 + |\mathrm{Ent}(\theta_k)|} \geq \epsilon.$$

- Ratio:

$$r_k = \frac{\eta(\tilde{\theta}_{k+1}) - \eta(\theta_k)}{L_{\theta_k}(\tilde{\theta}_{k+1}) - L_{\theta_k}(\theta_k)} \Longrightarrow r_k = \frac{\hat{\eta}(\tilde{\theta}_{k+1}) - \hat{\eta}(\theta_k)}{\hat{\sigma}_\eta(\theta_k) + \hat{L}_{\theta_k}(\tilde{\theta}_{k+1}) - \hat{L}_{\theta_k}(\theta_k)}.$$

- $\hat{\sigma}_\eta(\theta)$ is the empirical standard deviation of $\eta(\theta)$.
- Acceptance criteria: $\theta_{k+1} = \begin{cases} \tilde{\theta}_{k+1}, & r_k \geq \beta_0, \\ \theta_k, & \text{o.w.} \end{cases}$, with a small negative constant $\beta_0 < 0$.
- Mandatory acceptance: after several consecutive rejections, force to accept the best performed point among the past rejections.

Mujoco in Baselines

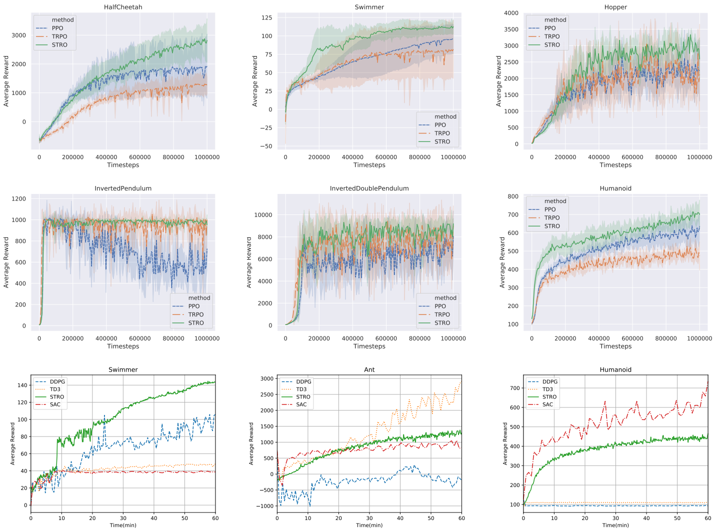


Figure: Training curves on Mujoco-v2 continuous control benchmarks.

Atari games

Table: Max Average Reward (100 episodes) $\pm$ standard deviation over 5 trails of 1e7 time steps.

Environment	PPO	TRPO	STRO
Pong	20±0	3±7	**20**±0
MsPacman	2125±322	1538±159	**2452**±487
Seaquest	1004±141	692±92	**1172**±346
Bowling	50±17	38±15	**105**±6
Freeway	30±0	28±3	**31**±0
PrivateEye	100±0	88±16	**100**±0

Many Thanks For Your Attention!

- 北大课程：大数据分析中的算法，华文慕课回放 http://bicmr.pku.edu.cn/~wenzw/bigdata2020.html
- 教材：刘浩洋, 户将, 李勇锋，文再文，最优化计算方法http://bicmr.pku.edu.cn/~wenzw/optbook.html
- Looking for Ph.D students and Postdoc
  Competitive salary as U.S and Europe
- http://bicmr.pku.edu.cn/~wenzw
- E-mail: wenzw@pku.edu.cn
- Office phone: 86-10-62744125