Technische Universität Braunschweig

Institut für Analysis und Algebra

Asymptotic Analysis of Unrolled Convex Optimization Algorithms

Christoph Brauer and Dirk Lorenz

Introduction

Task

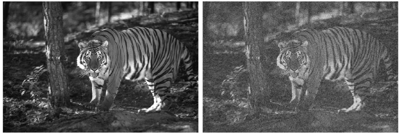
Recover ground truth $\boldsymbol{y} \in \mathbb{R}^n$
from noisy observation $\boldsymbol{x} \in \mathbb{R}^n$

use $\hat{\boldsymbol{K}}$

Bilevel Problem

$$\widehat{\boldsymbol{K}} \in \operatorname*{argmin}_{\boldsymbol{K}} \quad \frac{1}{m}\sum_{i=1}^{m} \ell(\boldsymbol{y}_i, \hat{\boldsymbol{y}}_i) \qquad \text{s.t.} \quad \forall i : \hat{\boldsymbol{y}}_i \in S(\boldsymbol{K}, \boldsymbol{x}_i)$$

via



y x

via

Convex Problem

reg. data fit

$$\hat{\boldsymbol{y}} \in S(\boldsymbol{K}, \boldsymbol{x}) := \operatorname*{argmin}_{\boldsymbol{u}} F(\boldsymbol{K}\boldsymbol{u}) + G(\boldsymbol{u} - \boldsymbol{x})$$

$K \in \mathbb{R}^{k \times n}$

learn $\boldsymbol{K}$
from

Training Data

$$\{(\boldsymbol{x}_1, \boldsymbol{y}_1), \ldots, (\boldsymbol{x}_m, \boldsymbol{y}_m)\}$$

Introduction

Bilevel Problem

$$\begin{aligned} \widehat{\boldsymbol{K}} \in \operatorname*{argmin}_{\boldsymbol{K}} \quad & \frac{1}{m}\sum_{i=1}^{m} \ell(\boldsymbol{y}_i, \hat{\boldsymbol{y}}_i) \\ \text{s.t.} \quad & \forall i : \hat{\boldsymbol{y}}_i \in S(\boldsymbol{K}, \mathbf{x}_i) \end{aligned}$$

unrolling

Convex Optimization Algorithm

$$A^1(\boldsymbol{K}, \boldsymbol{x}), \ldots, A^L(\boldsymbol{K}, \boldsymbol{x})$$

Approximate Bilevel Problem

$$\widehat{\boldsymbol{K}} \in \operatorname*{argmin}_{\boldsymbol{K}} \frac{1}{m}\sum_{i=1}^{m} \ell(\boldsymbol{y}_i, A^L(\boldsymbol{K}, \boldsymbol{x}_i))$$

Introduction

Approximate Bilevel Problem

$$\widehat{\boldsymbol{K}} \in \operatorname*{argmin}_{\boldsymbol{K}} \frac{1}{m} \sum_{i=1}^{m} \underline{\ell(\boldsymbol{y}_i, A^L(\boldsymbol{K}, \boldsymbol{x}_i))}$$

single

example

(x, y)

Approximate Gradient

$$\nabla_{\boldsymbol{K}}\, \ell(\boldsymbol{y}, A^L(\boldsymbol{K}, \boldsymbol{x})) \xrightarrow[\text{for } \underline{L \to \infty}]{\text{behavior}} ?$$

Gradient Computation

$F(Ku) + G(u - x)$

$\ell = L, \dots, 1$

$\ell(y, x^{[L]}) = \frac{1}{2}\|y - x^{[L]}\|^2$

Chambolle-Pock iteration

$$\mathbf{z}_D^{[\ell+1]} := \xi^{[\ell]} + \sigma \mathbf{K}\overline{\mathbf{x}}^{[\ell]} \qquad \xi^{[\ell+1]} := \mathrm{prox}_{\sigma F^*}(\mathbf{z}_D^{[\ell+1]})$$

$$\mathbf{z}_P^{[\ell+1]} := \mathbf{x}^{[\ell]} - \tau \mathbf{K}^\top \xi^{[\ell+1]} \qquad \mathbf{x}^{[\ell+1]} := \mathrm{prox}_{\tau G}(\mathbf{z}_P^{[\ell+1]} - \mathbf{x}) + \mathbf{x}$$

$$\overline{\mathbf{x}}^{[\ell+1]} := \mathbf{x}^{[\ell+1]} + \theta(\mathbf{x}^{[\ell+1]} - \mathbf{x}^{[\ell]})$$

$x^{[0]} = \bar{x}^{[0]} = x$

$\xi^{[0]} = 0$

$\ell = 0, \dots, L-1$

$A^L(u, x) = x^{[L]}$

$$\delta_{P/D}^{[\ell]} := \nabla_{\mathbf{z}_{P/D}^{[\ell]}} \ell(\mathbf{y}, \mathbf{x}^{[L]})$$

Backpropagated gradients

$$\delta_P^{[\ell]} = \mathrm{prox}'_{\tau G}(\mathbf{z}_P^{[\ell]} - \mathbf{x}) \odot (\delta_P^{[\ell+1]} + \sigma \mathbf{K}^\top \overline{\delta}_D^{[\ell+1]})$$

$$\delta_D^{[\ell]} = \mathrm{prox}'_{\sigma F^*}(\mathbf{z}_D^{[\ell]}) \odot (\delta_D^{[\ell+1]} - \tau \mathbf{K} \delta_P^{[\ell]})$$

$$\overline{\delta}_D^{[\ell]} = \delta_D^{[\ell]} + \theta(\delta_D^{[\ell]} - \delta_D^{[\ell+1]})$$

$\delta_D^{[L+1]} = \bar{\delta}_D^{[L+1]} = 0$

$\delta_P^{[L+1]} = y - x^{[L]}$

Parameter gradient

$$\nabla_{\mathbf{K}} \ell(\mathbf{y}, A^L(\mathbf{K}, \mathbf{x})) = \sum_{\ell=1}^{L} \sigma \delta_D^{[\ell]} \overline{\mathbf{x}}^{[\ell-1]\top} - \tau \xi^{[\ell]} \delta_P^{[\ell]\top}$$

Asymptotics of Backpropagated Gradients

Assumption

$$\left.\begin{array}{llll} \mathrm{prox}'_{\tau G}(\mathbf{z}_P^{[\ell]} - \mathbf{x}) & = & \mathrm{prox}'_{\tau G}(\mathbf{z}_P^{[\ell_0]} - \mathbf{x}) & \in \{0,1\}^n \\ \mathrm{prox}'_{\sigma F^*}(\mathbf{z}_D^{[\ell]}) & = & \mathrm{prox}'_{\sigma F^*}(\mathbf{z}_D^{[\ell_0]}) & \in \{0,1\}^k \end{array}\right\} \quad \begin{array}{l} \text{for some } \ell_0 \in \mathbb{N} \\ \text{and all } \ell \geq \ell_0 \end{array}$$

for fixed $\ell \geq \ell_0$

Result

$$\lim_{L\to\infty} \delta_P^{[\ell]} \in \ker(\boldsymbol{K}) \quad \text{and} \quad \lim_{L\to\infty} \delta_D^{[\ell]} \in \ker(\boldsymbol{K}^\top)$$

Proof

$$\lim_{L\to\infty} \delta_P^{[\ell]} \in \operatorname*{argmin}_{\delta_P} \text{ const. s.t. } U\delta_P = 0$$

$$\text{— " —} \quad \delta_D^{[\ell]} \in \operatorname*{argmin}_{\delta_D} \text{ const. s.t. } U^T\delta_D = 0$$

Asymptotics of Parameter Gradient

Assumptions

$$\mathbf{\Delta}_P := \lim_{L\to\infty} \sum_{\ell=1}^{L} \delta_P^{[\ell]} < \infty \qquad \mathbf{\Delta}_D := \lim_{L\to\infty} \sum_{\ell=1}^{L} \delta_D^{[\ell]} < \infty$$

$$\lim_{L\to\infty} \sum_{\ell=1}^{L} |\delta_P^{[\ell]}| < \infty \qquad \lim_{L\to\infty} \sum_{\ell=1}^{L} |\delta_D^{[\ell]}| < \infty$$

$$\downarrow$$

Result

$$\lim_{L\to\infty} \nabla_{\boldsymbol{K}}\, \ell(\boldsymbol{y}, \boldsymbol{x}^{[L]}) = \sigma \mathbf{\Delta}_D \boldsymbol{x}^{*\top} - \tau \boldsymbol{\xi}^* \mathbf{\Delta}_P^\top$$

Algorithmic Approach

Approximate Series

$$\Delta_P \approx \overset{\delta_P^{[L]}}{\underline{\delta_P^*}} := \operatorname{prox}'_{\tau G}(\overset{x^{[L]}}{\underline{\boldsymbol{x}^*}} - \tau \boldsymbol{K}^\top \boldsymbol{\xi}^* - \boldsymbol{x}) \odot (\boldsymbol{y} - \boldsymbol{x}^*)$$

$$\Delta_D \approx \underset{\delta_D^{[L]}}{\underline{\delta_D^*}} := \operatorname{prox}'_{\sigma F^*}(\underset{\xi^{[L]}}{\underline{\boldsymbol{\xi}^*}} + \sigma \boldsymbol{K}\boldsymbol{x}^*) \odot (-\tau \boldsymbol{K}\delta_P^*)$$

$\rightarrow$

Approximate Gradient

$$\lim_{L\to\infty} \nabla_{\boldsymbol{K}} \ell(\boldsymbol{y}, \boldsymbol{x}^{[L]}) \approx \underline{\sigma \delta_D^* \boldsymbol{x}^{*\top} - \tau \boldsymbol{\xi}^* \delta_P^{*\top}}$$

$\uparrow$

Fixed-Point Iteration

$$\tilde{\boldsymbol{\xi}} = \boxed{\operatorname{prox}_{\sigma F^*}(\boldsymbol{\xi}^* + \sigma \boldsymbol{K}\boldsymbol{x}^*) = x^*}$$

$$\tilde{\boldsymbol{x}} = \boxed{\operatorname{prox}_{\tau G}(\boldsymbol{x}^* - \tau \boldsymbol{K}^\top \underset{\xi^*}{\tilde{\boldsymbol{\xi}}}) = \xi^*}$$

$\rightarrow$

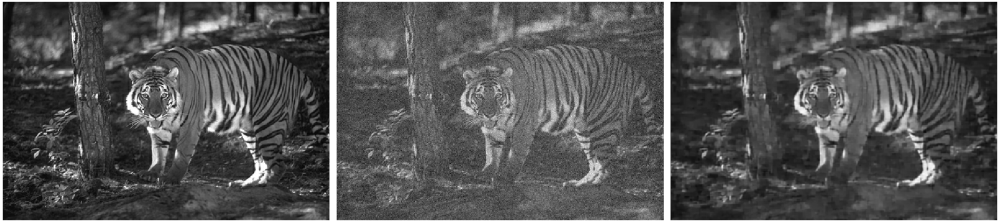
Fixed-Point Gradient

$$\underline{\nabla_{\boldsymbol{K}} \ell(\boldsymbol{y}, \tilde{\boldsymbol{x}})}$$

Numerical Example

$F = \|\cdot\|_1$ $\quad G = I$ $\quad 48 \quad 7 \times 7$

$$\hat{\boldsymbol{y}} \in \underset{\boldsymbol{u}}{\operatorname{argmin}} \|\boldsymbol{K}\boldsymbol{u}\|_1 \quad \text{s.t. } \|\boldsymbol{u} - \boldsymbol{x}\|_2 \leq \sigma \sqrt{\#\text{pixels}}$$



y $\qquad x$ $\qquad \hat{y}$

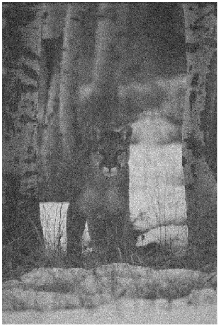
Numerical Example

$$\hat{\boldsymbol{y}} \quad \in \quad \underset{\boldsymbol{u}}{\operatorname{argmin}} \, \|\boldsymbol{K}\boldsymbol{u}\|_1 \quad \text{s.t. } \|\boldsymbol{u} - \boldsymbol{x}\|_2 \leq \sigma\sqrt{\#\text{pixels}}$$



Numerical Example

$$\hat{\boldsymbol{y}} \quad \in \quad \underset{\boldsymbol{u}}{\operatorname{argmin}} \, \|\boldsymbol{K}\boldsymbol{u}\|_1 \quad \text{s.t. } \|\boldsymbol{u}-\boldsymbol{x}\|_2 \leq \sigma\sqrt{\#\text{pixels}}$$



Numerical Example

$$\hat{\boldsymbol{y}} \in \underset{\boldsymbol{u}}{\operatorname{argmin}} \|\boldsymbol{K}\boldsymbol{u}\|_1 \quad \text{s.t. } \|\boldsymbol{u}-\boldsymbol{x}\|_2 \leq \sigma\sqrt{\#\text{pixels}}$$

Summary

- Asymptotic analysis of gradients in parameterized energy minimization models
- Gradient limit depends only on optimal solutions and not on intermediate iterates
- Approximating backpropagated gradients yields tractable gradient computation
- Interpretation in terms of fixed-point equation
- Application to image denoising

Thank you!