A Hamilton-Jacobi-based proximal operator

Stanley Osher[1], Howard Heaton[2], Samy Wu Fung[3]

[1]UCLA, [2]Typal Academy, [3]Colorado School of Mines

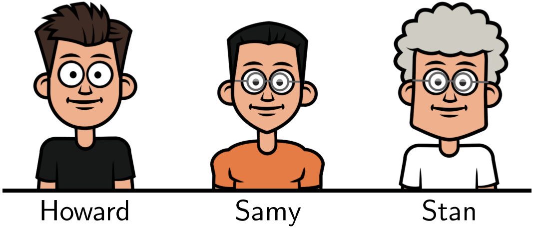
Overview

Scope: Optimization with proximals when we do not have exact formulas

Problem: For a continuous and weakly convex function $f : \mathbb{R}^n \to \mathbb{R}$ and $t > 0$, estimate

$$\operatorname{prox}_{tf}(x) = \operatorname*{argmin}_{y} f(y) + \frac{1}{2t}\|y - x\|^2,$$

only *using evaluations of* f.



Key Ingredients

- Moreau Envelope
- Hamilton-Jacobi PDEs
- Cole-Hopf Transformation
- Monte Carlo Sampling

Gradients of Moreau Envelopes

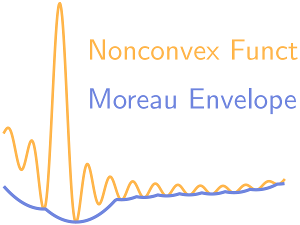
Given a function $f : \mathbb{R}^n \to \mathbb{R}$ and time $t > 0$, the Moreau envelope u is

$$u(x, t) \triangleq \min_y f(y) + \frac{1}{2t}\|y - x\|^2.$$

When it exists, the gradient is

$$\nabla u(x, t) = \frac{1}{t}(x - \mathrm{prox}_{tf}(x))$$

$$\implies \quad \mathrm{prox}_{tf}(x) = x - t\nabla u(x, t).$$



Nonconvex Function f

Moreau Envelope u

Hamilton-Jacobi PDEs

The envelope u solves

$$\begin{cases} u_t + \frac{1}{2}\|\nabla u\|^2 = 0 & \text{in } \mathbb{R}^n \times (0,\infty) \\ u = f & \text{on } \mathbb{R}^n \times \{t = 0\}. \end{cases}$$

For small $\delta > 0$, we can approximate u using the viscous PDE

$$\begin{cases} u^\delta_t + \frac{1}{2}\|\nabla u^\delta\|^2 = \frac{\delta}{2}\Delta u^\delta & \text{in } \mathbb{R}^n \times (0,\infty) \\ u^\delta = f & \text{on } \mathbb{R}^n \times \{t = 0\}. \end{cases}$$

Cole-Hopf Transformation

Using an idea from Cole and Hopf, we use the change of variables

$$v^\delta = \exp(-u^\delta/\delta),$$

for which v^δ solves the heat equation

$$\begin{cases} v_t^\delta - \frac{\delta}{2}\Delta v^\delta = 0 & \text{in } \mathbb{R}^n \times (0,\infty) \\ v^\delta = \exp(-f/\delta) & \text{on } \mathbb{R}^n \times \{t=0\}. \end{cases}$$

Monte Carlo Sampling

For a heat kernel $\Phi_{\delta t}$, the heat equation solution v^δ can be written as

$$v^\delta(x,t) = \Big(\Phi_{\delta t} * \exp(-f/\delta)\Big)(x) = \mathbb{E}_{y\sim\mathcal{N}(x,\delta t)}\left[\exp\left(-f(y)/\delta\right)\right],$$

with the expectation over a normal distribution with variance δt and mean x. Then

$$\nabla v^\delta(x,t) = -\frac{1}{\delta t}\cdot\mathbb{E}_{y\sim\mathcal{N}(x,\delta t)}\left[(x-y)\exp\left(-f(y)/\delta\right)\right],$$

which implies

$$\nabla u^\delta(x,t) = \frac{1}{t}\cdot\left(x - \frac{\mathbb{E}_{y\sim\mathcal{N}(x,\delta t)}\left[y\cdot\exp\left(-f(y)/\delta\right)\right]}{\mathbb{E}_{y\sim\mathcal{N}(x,\delta t)}\left[\exp\left(-f(y)/\delta\right)\right]}\right).$$

Key Result

Using the gradient of the Moreau envelope and our approximation,

$$\begin{aligned}\mathrm{prox}_{tf}(x) &= x - t\nabla u(x,t)\\ &\approx x - t\nabla u^{\delta}(x,t)\\ &= \frac{\mathbb{E}_{y\sim\mathcal{N}(x,\delta t)}\left[y\cdot\exp\left(-f(y)/\delta\right)\right]}{\mathbb{E}_{y\sim\mathcal{N}(x,\delta t)}\left[\exp\left(-f(y)/\delta\right)\right]}.\end{aligned}$$

(Informal) Theorem: If t is sufficiently small and $u(x,t) \geq 0$, then

$$\lim_{\delta\to 0^+} \frac{\mathbb{E}_{y\sim\mathcal{N}(x,\delta t)}\left[y\cdot\exp\left(-f(y)/\delta\right)\right]}{\mathbb{E}_{y\sim\mathcal{N}(x,\delta t)}\left[\exp\left(-f(y)/\delta\right)\right]} = \mathrm{prox}_{tf}(x).$$

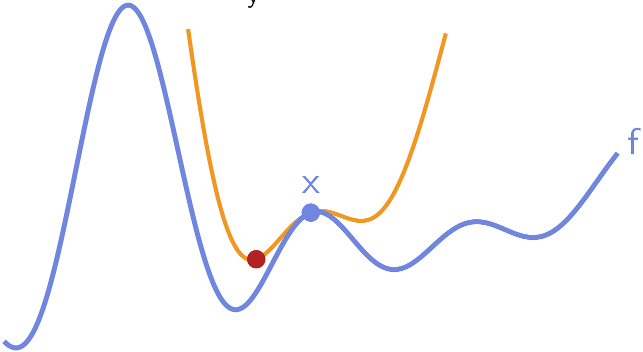
Implementation

Algorithm 1 HJ-Prox – Approximation of Proximal Operator

```
HJ-Prox($x,\ t;\ f,\ \delta,\ N,\ \alpha,\ \varepsilon$) :
    for $i \in [N]$:
        Sample $y^i \sim \mathcal{N}(x, \delta t/\alpha)$
        $z_i \leftarrow f(y^i)$
    if $\exp\left(-\alpha z_i/\delta\right) \leq \varepsilon$ for large proportion of samples:
        return HJ-Prox($x,\ t;\ f,\ \delta,\ N,\ \alpha/2,\ \varepsilon$)
    prox $\leftarrow$ softmax$(-\alpha z/\delta)^\top[y^1 \cdots y^N]$
    return prox
```

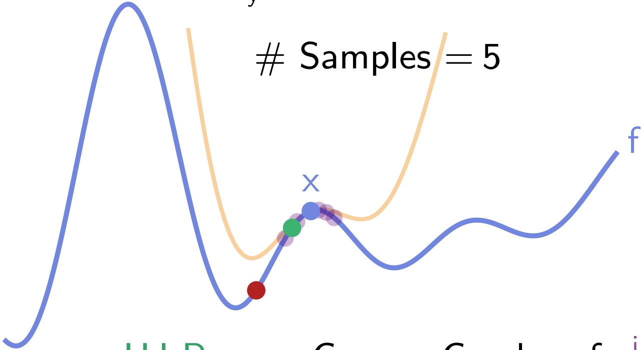
Proximal Computation Example

$$\mathrm{prox}_{tf}(x) = \operatorname*{argmin}_{y} f(y) + \frac{1}{2t}\|y - x\|^2$$



Proximal Computation Example

$$\mathrm{prox}_{tf}(x) = \underset{y}{\mathrm{argmin}}\ f(y) + \frac{1}{2t}\|y - x\|^2$$



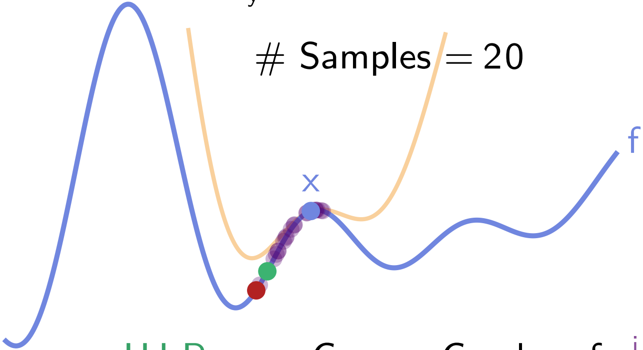
$\#$ Samples $= 5$

HJ-Prox $=$ Convex Combo of y^i

y^i weight $= \mathrm{softmax}(-f(y)/\delta)$

Proximal Computation Example

$$\mathrm{prox}_{tf}(x) = \underset{y}{\mathrm{argmin}}\ f(y) + \frac{1}{2t}\|y - x\|^2$$



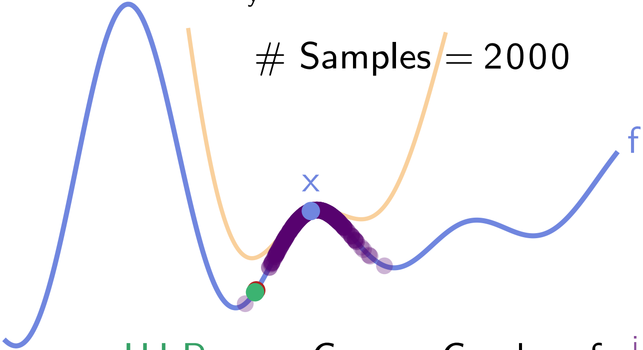
$\#$ Samples $= 20$

HJ-Prox $=$ Convex Combo of y^i

y^i weight $= \mathrm{softmax}(-f(y)/\delta)$

Proximal Computation Example

$$\mathrm{prox}_{tf}(x) = \operatorname*{argmin}_{y}\ f(y) + \frac{1}{2t}\|y - x\|^2$$



HJ-Prox = Convex Combo of y^i

y^i weight = $\mathrm{softmax}(-f(y)/\delta)$

Numerical Example in Optimization Algorithm

Consider a constrained minimization problem where objective values f can only be accessed via a noisy oracle $\mathcal{O}$, *i.e.*

$$\min_{x\in\mathbb{R}^{1000}} \mathbb{E}[\mathcal{O}(x)] \quad \text{s.t.} \quad Ax = b,$$

where the expectation is over the noise. We assume A is large and under-determined. Using HJ-Prox, we solve with the linearized method of multipliers via updates

$$\begin{aligned} x^{k+1} &= \operatorname{prox}_{t\mathcal{O}}\left(x^k - tA^\top(u^k + \lambda(Ax^k - b))\right) \\ u^{k+1} &= u^k + \lambda(Ax^{k+1} - b). \end{aligned}$$

Relative Error Plots with HJ-Prox

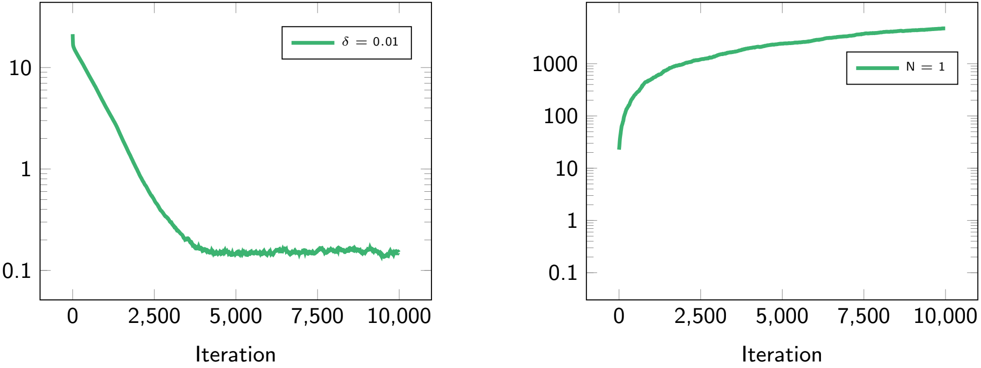


Figure 1: Plots of linearized method of multipliers using HJ-Prox estimates for each proximal. Gradient descent on constraint violation is provided for reference and does not use the oracle $\mathcal{O}$.

Relative Error Plots with HJ-Prox

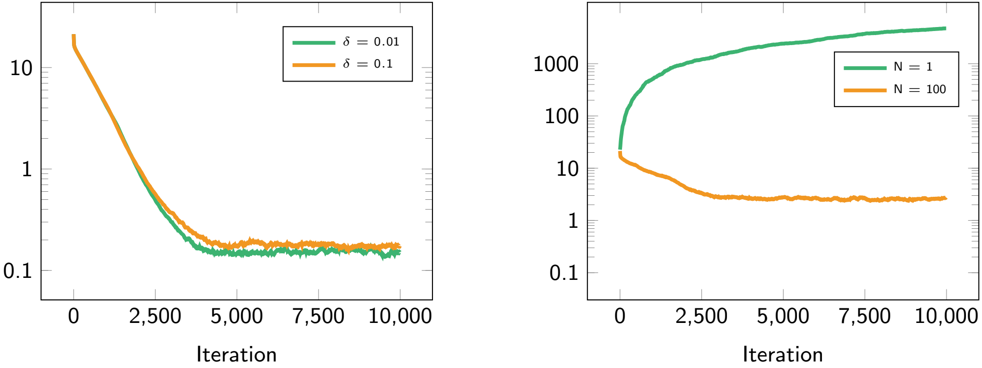


Figure 2: Plots of linearized method of multipliers using HJ-Prox estimates for each proximal. Gradient descent on constraint violation is provided for reference and does not use the oracle $\mathcal{O}$.

Relative Error Plots with HJ-Prox

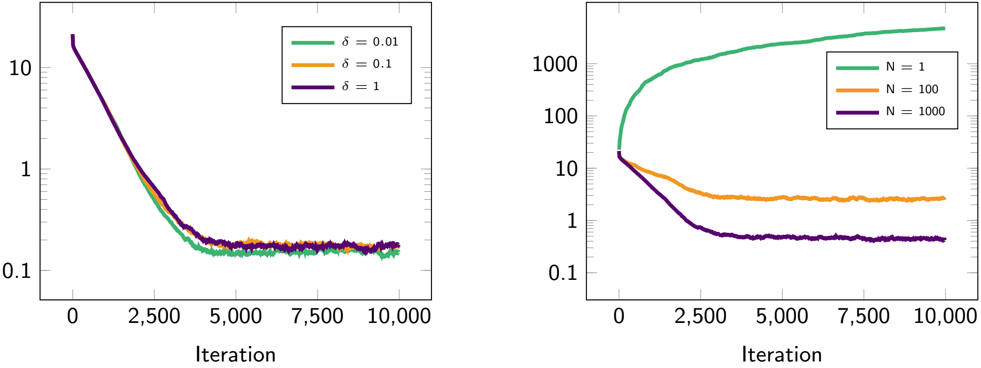


Figure 3: Plots of linearized method of multipliers using HJ-Prox estimates for each proximal. Gradient descent on constraint violation is provided for reference and does not use the oracle $\mathcal{O}$.

Relative Error Plots with HJ-Prox

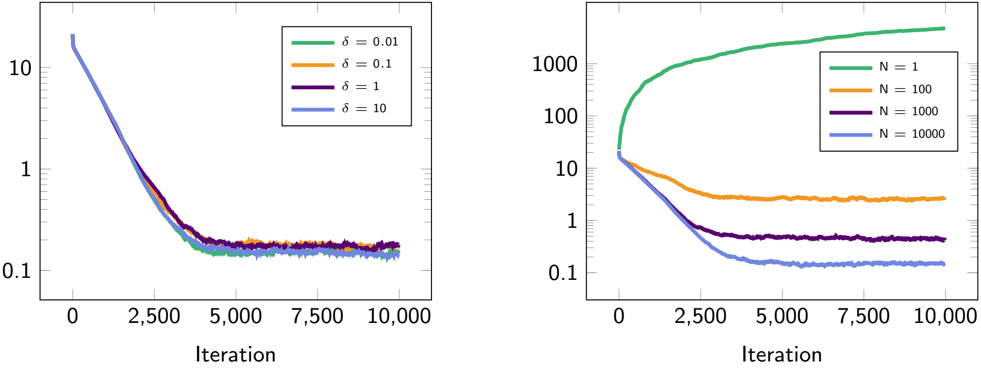


Figure 4: Plots of linearized method of multipliers using HJ-Prox estimates for each proximal. Gradient descent on constraint violation is provided for reference and does not use the oracle $\mathcal{O}$.

Relative Error Plots with HJ-Prox

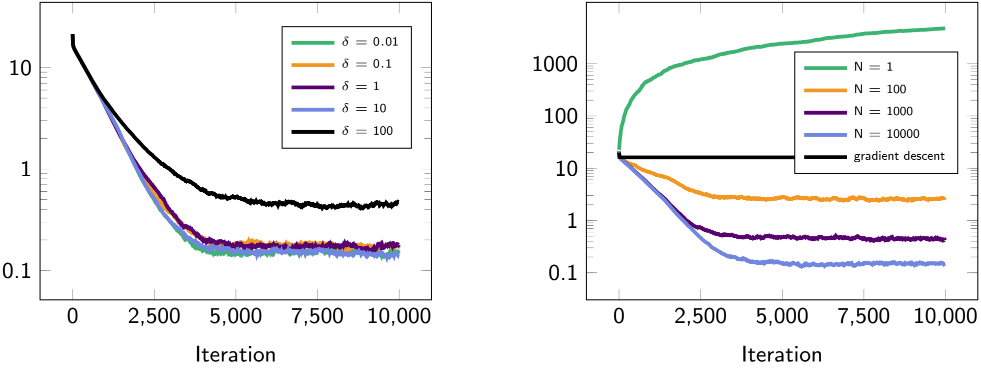


Figure 5: Plots of linearized method of multipliers using HJ-Prox estimates for each proximal. Gradient descent on constraint violation is provided for reference and does not use the oracle $\mathcal{O}$.

Summary

- HJ-Prox gives a simple zeroth-order approximation to proximals
- The parameter δ smooths approximations (potentially helpful for denoising)
- HJ-Prox can be embedded inside optimization algorithms
  (*e.g.* proximal gradient, Douglas Rachford, ADMM, PDHG)

Preprint: arxiv.org/abs/2211.12997

Reprint: https://www.pnas.org/doi/10.1073/pnas.2220469120