A High-Temperature Perspective on the Zeroth-Order Proximal Operator

Emanuele Naldi, **Hippolyte Labarrière**, Cesare Molinari, Silvia Villa

EUROPT 2026
Linz, Austria

July 9, 2026

Background

Zeroth-Order Optimization

$$\min_{x \in \mathbb{R}^d} f(x)$$

BUT no first-order information available (no ∇f, no ∂f)!

Background

Zeroth-Order Optimization

$$\min_{x \in \mathbb{R}^d} f(x)$$

BUT no first-order information available (no ∇f, no ∂f)!

Applications:

- Adversarial attacks,
- Robotics,
- Simulation-based optimization...

Background

The proximal operator [Moreau '65]: for a (weakly) convex f,

$$\operatorname{prox}_{\lambda f}(x) = \operatorname*{arg\,min}_{y \in \mathbb{R}^d} \left\{ f(y) + \frac{1}{2\lambda} \|y - x\|^2 \right\}$$

Powerful tool for convex optimization:

- $\operatorname{prox}_{\lambda f} = (I + \lambda \partial f)^{-1}$
- $\operatorname{arg\,min}_{\mathbb{R}^d} f \in \mathsf{Fix}(\operatorname{prox}_{\lambda f})$
- $f(\operatorname{prox}_{\lambda f}(x)) \leq f(x)$

Background

The proximal operator [Moreau '65]:

$$\operatorname{prox}_{\lambda f}(x) = \underset{y \in \mathbb{R}^d}{\arg\min} \left\{ f(y) + \frac{1}{2\lambda} \|y - x\|^2 \right\}$$

$\rightarrow$ Building brick of **proximal methods**:

- **Proximal Point Algorithm [Güler '91]**: $x^{k+1} = \operatorname{prox}_{\lambda f}(x^k)$
- **Proximal Gradient (Forward-Backward) [Lions, Mercier '79, Combettes, Wajs '05]**:

$$x^{k+1} = \operatorname{prox}_{\lambda h}(x^k - \lambda \nabla f(x^k))$$

- **Other Splitting Methods**: Douglas-Rachford, Primal-Dual...

Background

The proximal operator [Moreau ’65]:

$$\operatorname{prox}_{\lambda f}(x) = \arg\min_{y \in \mathbb{R}^d} \left\{ f(y) + \frac{1}{2\lambda} \|y - x\|^2 \right\}$$

Powerful **BUT not always explicit!**

Approximating the prox

Relating the prox to Hamilton-Jacobi (HJ)

$$\begin{cases} \partial_t u(x,t) + \frac{1}{2} \left\| \nabla u(x,t) \right\|^2 = 0 \\ u(x,0) = f(x) \end{cases}$$

Approximating the prox

Relating the prox to Hamilton-Jacobi (HJ)

$$\begin{cases} \partial_t u(x,t) + \frac{1}{2} \left\| \nabla u(x,t) \right\|^2 = 0 \\ u(x,0) = f(x) \end{cases}$$

The solution of (HJ) is the **Moreau envelope**:

$$u(x,t) = \min_{y \in \mathbb{R}^d} \left\{ f(y) + \frac{1}{2t} \|y - x\|^2 \right\}$$

Approximating the prox

Relating the prox to Hamilton-Jacobi (HJ)

$$\begin{cases} \partial_t u(x,t) + \frac{1}{2}\left\|\nabla u(x,t)\right\|^2 = 0 \\ u(x,0) = f(x) \end{cases}$$

The solution of (HJ) is the **Moreau envelope**, which can be related to the prox!

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

Approximating the prox

Add a perturbation (Hamilton-Jacobi-Bellman) [Osher et al. '23]

$$\begin{cases} \partial_t u^\delta(x,t) + \frac{1}{2}\left\|\nabla u^\delta(x,t)\right\|^2 = \frac{\delta}{2}\Delta u^\delta \\ u^\delta(x,0) = f(x) \end{cases}$$

Approximating the prox

Add a perturbation (Hamilton-Jacobi-Bellman) [Osher et al. '23]

$$\begin{cases} \partial_t u^\delta(x,t) + \frac{1}{2}\left\|\nabla u^\delta(x,t)\right\|^2 = \frac{\delta}{2}\Delta u^\delta \\ u^\delta(x,0) = f(x) \end{cases}$$

The solution of (HJB) is explicit!

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

Approximating the prox

Add a perturbation (Hamilton-Jacobi-Bellman) [Osher et al. '23]

$$\begin{cases} \partial_t u^{\delta}(x,t) + \frac{1}{2}\left\|\nabla u^{\delta}(x,t)\right\|^2 = \frac{\delta}{2}\Delta u^{\delta} \\ u^{\delta}(x,0) = f(x) \end{cases}$$

The solution of (HJB) is explicit , so let's define

$$\mathrm{zprox}_{t,f}^{\delta}(x) := x - t\nabla_x u^{\delta}(x,t)$$

Approximating the prox

Add a perturbation (Hamilton-Jacobi-Bellman) [Osher et al. '23]

$$\begin{cases} \partial_t u^\delta(x,t) + \frac{1}{2}\left\|\nabla u^\delta(x,t)\right\|^2 = \frac{\delta}{2}\Delta u^\delta \\ u^\delta(x,0) = f(x) \end{cases}$$

The solution of (HJB) is explicit , so let's define

$$\mathrm{zprox}_{t,f}^{\delta}(x) = \frac{\mathbb{E}_{y\sim\mathcal{N}(x,\delta t)}\left[y\exp\left(-\frac{f(y)}{\delta}\right)\right]}{\mathbb{E}_{y\sim\mathcal{N}(x,\delta t)}\left[\exp\left(-\frac{f(y)}{\delta}\right)\right]}$$

A zeroth order approximation of the prox!

Zeroth Order Proximal Operator

ZOPO: A Zeroth Order approximation of the proximal operator...

$$\mathrm{zprox}^{\delta}_{\lambda,f}(x) = \frac{\mathbb{E}_{y\sim\mathcal{N}(x,\delta\lambda)}\left[y\exp\left(-\frac{f(y)}{\delta}\right)\right]}{\mathbb{E}_{y\sim\mathcal{N}(x,\delta\lambda)}\left[\exp\left(-\frac{f(y)}{\delta}\right)\right]}$$

Zeroth Order Proximal Operator

ZOPO: A Zeroth Order approximation of the proximal operator...

$$\mathrm{zprox}^{\delta}_{\lambda,f}(x) = \frac{\mathbb{E}_{y\sim\mathcal{N}(x,\delta\lambda)}\left[y\exp\left(-\frac{f(y)}{\delta}\right)\right]}{\mathbb{E}_{y\sim\mathcal{N}(x,\delta\lambda)}\left[\exp\left(-\frac{f(y)}{\delta}\right)\right]}$$

...which recovers the prox [Osher et al. '23]: under weak convexity and smoothness/Lipschitzness

$$\lim_{\delta\to 0}\mathrm{zprox}^{\delta}_{\lambda,f}(x) = \mathrm{prox}_{\lambda f}(x)$$

Zeroth Order Proximal Operator

ZOPO: A Zeroth Order approximation of the proximal operator...

$$\mathrm{zprox}^{\delta}_{\lambda,f}(x) = \frac{\mathbb{E}_{y\sim\mathcal{N}(x,\delta\lambda)}\left[y \exp\left(-\frac{f(y)}{\delta}\right)\right]}{\mathbb{E}_{y\sim\mathcal{N}(x,\delta\lambda)}\left[\exp\left(-\frac{f(y)}{\delta}\right)\right]}$$

...which recovers the prox [Osher et al. '23]: under weak convexity and smoothness/Lipschitzness

$$\lim_{\delta\to 0} \mathrm{zprox}^{\delta}_{\lambda,f}(x) = \mathrm{prox}_{\lambda f}(x)$$

Still not explicit!

Sampling for ZOPO

ZOPO can be estimated[a] **by**

$$\widehat{\mathrm{zprox}}_{\lambda,f}^{\delta,N}(x) = \frac{\sum_{i=1}^{N} y_i \exp\left(-\frac{f(y_i)}{\delta}\right)}{\sum_{i=1}^{N} \exp\left(-\frac{f(y_i)}{\delta}\right)}, \quad y_i \sim \mathcal{N}(x, \lambda\delta)$$

$\rightarrow$ **Explicit** and **Zeroth Order**!

[a] see also Tensor Train approximation [Zhang et al. '24]

Sampling for ZOPO

ZOPO can be estimated by

$$\widehat{\mathrm{zprox}}_{\lambda,f}^{\delta,N}(x) = \frac{\sum_{i=1}^{N} y_i \exp\left(-\frac{f(y_i)}{\delta}\right)}{\sum_{i=1}^{N} \exp\left(-\frac{f(y_i)}{\delta}\right)}, \quad y_i \sim \mathcal{N}(x, \lambda\delta)$$

$\rightarrow$ **Explicit** and **Zeroth Order**!

Common strategy in the literature [Zhang et al. '24, Luo et al. '25, Di et al. '26]:

$\rightarrow$ Build a scheme based on $\widehat{\mathrm{zprox}}$ (Proximal Point/Splitting)

Sampling for ZOPO

ZOPO can be estimated by

$$\widehat{\mathrm{zprox}}_{\lambda,f}^{\delta,N}(x) = \frac{\sum_{i=1}^{N} y_i \exp\left(-\frac{f(y_i)}{\delta}\right)}{\sum_{i=1}^{N} \exp\left(-\frac{f(y_i)}{\delta}\right)}, \quad y_i \sim \mathcal{N}(x, \lambda\delta)$$

$\rightarrow$ **Explicit** and **Zeroth Order**!

Common strategy in the literature [Zhang et al. '24, Luo et al. '25, Di et al. '26]:

$\rightarrow$ Build a scheme based on $\widehat{\mathrm{zprox}}$ (Proximal Point/Splitting)

$\rightarrow$ Ensure δ goes to 0

Sampling for ZOPO

ZOPO can be estimated by

$$\widehat{\mathrm{zprox}}_{\lambda,f}^{\delta,N}(x) = \frac{\sum_{i=1}^{N} y_i \exp\left(-\frac{f(y_i)}{\delta}\right)}{\sum_{i=1}^{N} \exp\left(-\frac{f(y_i)}{\delta}\right)}, \quad y_i \sim \mathcal{N}(x, \lambda\delta)$$

$\rightarrow$ **Explicit** and **Zeroth Order**!

Common strategy in the literature [Zhang et al. '24, Luo et al. '25, Di et al. '26]:

$\rightarrow$ Build a scheme based on $\widehat{\mathrm{zprox}}$ (Proximal Point/Splitting)

$\rightarrow$ Ensure δ goes to 0

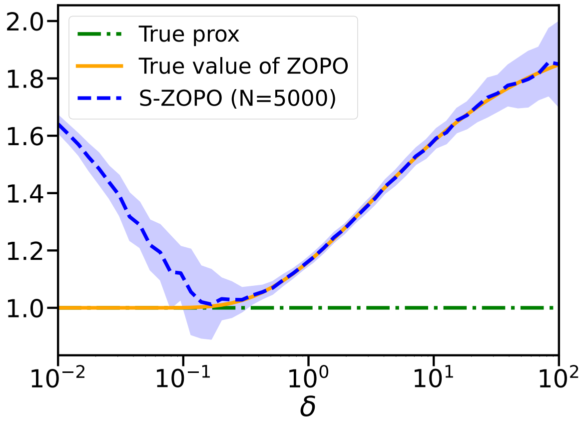
$\rightarrow$ Treat $\widehat{\mathrm{zprox}}$ as an inexact prox

The low-temperature regime

What happens for small δ?

The low-temperature regime

What happens for small δ? $\mathrm{zprox}^{\delta}_{\lambda,f}$ becomes harder to estimate!

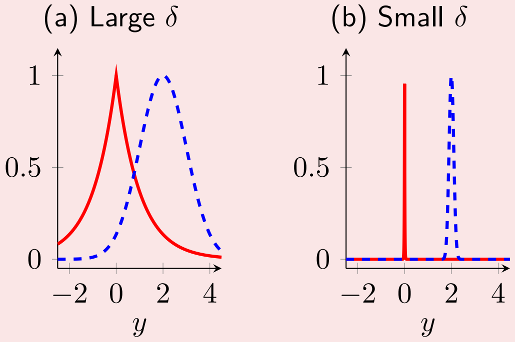


The low-temperature regime

What happens for small δ? $\mathrm{zprox}^{\delta}_{\lambda,f}$ becomes harder to estimate!

Key quantity

$$\mathbb{E}_{y\sim\mathcal{N}(x,\lambda\delta)}\left[\exp\left(-\frac{f(y)}{\delta}\right)\right]$$



The low-temperature regime

What happens for small δ? It becomes a **Random search operator!**

$$\widehat{\text{zprox}}_{\lambda,f}^{\delta,N}(x) = \frac{\sum_{i=1}^{N} y_i \exp\left(-\frac{f(y_i)}{\delta}\right)}{\sum_{i=1}^{N} \exp\left(-\frac{f(y_i)}{\delta}\right)} \approx \arg\min_{y_i} f(y_i)$$

The low-temperature regime

What happens for small δ? It becomes a **Random search operator!**

$$\widehat{\mathrm{zprox}}_{\lambda,f}^{\delta,N}(x) = \frac{\sum_{i=1}^{N} y_i \exp\left(-\frac{f(y_i)}{\delta}\right)}{\sum_{i=1}^{N} \exp\left(-\frac{f(y_i)}{\delta}\right)} \approx \arg\min_{y_i} f(y_i)$$

$\rightarrow$ Related to the **Effective Sample Size** of the estimator!

Our work

What happens for a fixed δ?

Our work

What happens for a fixed δ?

ZOPO is the **proximal operator of another potential!**

Why?

Our work

What happens for a fixed δ?

ZOPO is the **proximal operator of another potential!**

Why?

$$\mathrm{zprox}^{\delta}_{\lambda,f}(x) = \frac{\mathbb{E}_{y\sim\mathcal{N}(x,\delta\lambda)}\left[y\exp\left(-\frac{f(y)}{\delta}\right)\right]}{\mathbb{E}_{y\sim\mathcal{N}(x,\delta\lambda)}\left[\exp\left(-\frac{f(y)}{\delta}\right)\right]} = \mathbb{E}_{y\sim p(y|x)}[y]$$

where $p(y|x) \propto \exp(-\frac{1}{\delta}(f(y) + \frac{1}{\lambda}\|x-y\|^2))$.

$\rightarrow$ it is an **MMSE estimator**

Our work

What happens for a fixed δ?

ZOPO is the **proximal operator of another potential!**

Why?

$$\mathrm{zprox}^{\delta}_{\lambda,f}(x) = \frac{\mathbb{E}_{y\sim\mathcal{N}(x,\delta\lambda)}\left[y\exp\left(-\frac{f(y)}{\delta}\right)\right]}{\mathbb{E}_{y\sim\mathcal{N}(x,\delta\lambda)}\left[\exp\left(-\frac{f(y)}{\delta}\right)\right]} = \mathbb{E}_{y\sim p(y|x)}[y]$$

where $p(y|x) \propto \exp(-\frac{1}{\delta}(f(y) + \frac{1}{\lambda}\|x-y\|^2))$.

$\rightarrow$ it is an **MMSE estimator**

$\rightarrow$ it is a prox![Gribonval '11, Darbon et al. '21]

Our work

ZOPO is a PO!

- the underlying potential is **explicit**:

$$H_{\lambda,\delta}(x) = -\frac{1}{2\lambda}\left\|(\mathrm{zprox}_{\lambda,f}^{\delta})^{-1}(x) - x\right\|^2 + f^{\lambda,\delta}\Big((\mathrm{zprox}_{\lambda,f}^{\delta})^{-1}(x)\Big)$$

- Zeroth Order PPA **is a PPA on** $H_{\lambda,\delta}$! **No error propagation!**

Our work

ZOPO is a PO!

- the underlying potential is **explicit**
- Zeroth Order PPA **is a PPA on** $H_{\lambda,\delta}$! **No error propagation!**

Theorem [Naldi, L., Molinari, Villa '26] Under minimal assumptions, $\left(x^k\right)_k$ such that

$$x^{k+1} = \mathrm{zprox}^{\delta}_{\lambda,f}(x^k)$$

converges to $x^*_{\lambda,\delta}$ a fixed point $\mathrm{zprox}^{\delta}_{\lambda,f}$ which is a critical point of the viscous Moreau envelope $f^{\lambda,\delta}$. Also, $\sum_k \left\|\nabla f^{\lambda,\delta}(x^k)\right\| < +\infty$.

Our work

What does it mean to be a critical point of $f^{\lambda,\delta}$?

Our work

What does it mean to be a critical point of $f^{\lambda,\delta}$? We can relate $x^*_{\lambda,\delta}$ to f!
E.g. if f is L-Lipschitz and $\lambda < \frac{1}{L}$ (see also [Lauga, Vaiter '26]):

$$\|\nabla f(x^*_{\lambda,\delta})\| \leq L\sqrt{\frac{d\lambda\delta}{1-\lambda L}}$$

Our work

What does it mean to be a critical point of $f^{\lambda,\delta}$? We can relate $x^*_{\lambda,\delta}$ to f!
E.g. if f is L-Lipschitz and $\lambda < \frac{1}{L}$ (see also [Lauga, Vaiter '26]):

$$\|\nabla f(x^*_{\lambda,\delta})\| \leq L\sqrt{\frac{d\lambda\delta}{1-\lambda L}}$$

What else do we have?

- **sharper rates** and **characterization of the critical points** in the convex case,
- insights on the effect of λ: **large values convexify** $f^{\lambda,\delta}$**!**,
- **convergence proof** for the **sampled algorithm** in the convex case.

Conclusion

Take-away message:

- $\mathrm{prox}_{\lambda f}$ can be approximated by a zeroth order object: $\mathrm{zprox}^{\delta}_{\lambda,f}$
- **Low temperature regime** $(\delta \to 0)$: better approximation, worse sampling.
- **High temperature regime**: still a proximal operator, convergence guarantees, better sampling.

Conclusion

Take-away message:

- $\text{prox}_{\lambda f}$ can be approximated by a zeroth order object: $\text{zprox}^{\delta}_{\lambda,f}$
- **Low temperature regime** ($\delta \to 0$): better approximation, worse sampling.
- **High temperature regime**: still a proximal operator, convergence guarantees, better sampling.

Perspectives:

$\to$ Leverage theory to develop **efficient ZOPO algorithms** (Graduated optimization, adaptive parameters, better sampling)

$\to$ Derive **tighter results** for the sampled algorithm

Thank you for your attention!

Questions?

Related paper:

Naldi, E., Labarrière, H., Molinari, C., Villa, S., *Convergence of zeroth-order proximal point algorithms in the high-temperature regime*, arXiv:2605.11929, 2026.

My Website:

`https://hippolytelbrrr.github.io/pages/index_eng.html`

References

- Moreau, *Proximité et dualité dans un espace hilbertien*, Bulletin de la Société Mathématique de France, 1965.
- Güler, *On the convergence of the proximal point algorithm for convex minimization*, SIOPT, 1991.
- Lions, Mercier, *Splitting algorithms for the sum of two nonlinear operators*, SIAM Journal on Numerical Analysis, 1979.
- Combettes, Wajs, *Signal recovery by proximal forward-backward splitting*, Multiscale Modeling & Simulation, 2005.
- Osher, Heaton, Fung, *A Hamilton-Jacobi-based proximal operator*, PNAS, 2023.
- Heaton, Fung, Osher, *Global Solutions to Nonconvex Problems by Evolution of Hamilton-Jacobi PDEs*, Communications on Applied Mathematics and Computation, 2023.
- Zhang, Han, Chow, Osher, Schaeffer,*Inexact proximal point algorithms for zeroth-order global optimization*, 2024.
- Luo, Xu, Rabitz, *Asymptotic proximal point methods for global optimization*, 2025.
- Di, Chi, Fung, *Operator splitting with Hamilton-Jacobi-based proximals*, 2026.
- Gribonval, *Should penalized least squares regression be interpreted as maximum a posteriori estimation?*, IEEE Transactions on Signal Processing, 2011.
- Darbon, Langlois, Meng, *Connecting Hamilton-Jacobi partial differential equations with maximum a posteriori and posterior mean estimators for some non-convex priors*, 2021.
- Lauga, Vaiter, *Characterizations of inexact proximal operators*, 2026.