

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

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

STOP Workshop : STructured OPtimization for inverse and learning problems
Lyon, France

September 30, 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 ,

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

Powerful tool for convex optimization:

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

Background

The proximal operator [Moreau '65]:

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

→ Building brick of **proximal methods**:

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

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

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

Background

The proximal operator [Moreau '65]:

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

Powerful **BUT** rarely explicit!

Approximating the prox

Relating the prox to Hamilton-Jacobi (HJ)

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

Approximating the prox

Relating the prox to Hamilton-Jacobi (HJ)

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

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)

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

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

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

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]

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

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]

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

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

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

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

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

A zeroth order approximation of the prox!

Zeroth Order Proximal Operator

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

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

Zeroth Order Proximal Operator

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

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

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

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

Zeroth Order Proximal Operator

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

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

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

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

Still not explicit!

Sampling for ZOPO

ZOPO can be estimated^a by

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

→ **Explicit** and **Zeroth Order**!

^asee also Tensor Train approximation [Zhang et al. '24]

Sampling for ZOPO

ZOPO can be estimated by

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

→ **Explicit** and **Zeroth Order**!

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

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

Sampling for ZOPO

ZOPO can be estimated by

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

→ **Explicit** and **Zeroth Order**!

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

- Build a scheme based on $\widehat{\text{zprox}}$ (Proximal Point/Splitting)
- Ensure δ goes to 0

Sampling for ZOPO

ZOPO can be estimated by

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

→ **Explicit** and **Zeroth Order**!

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

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

→ Ensure δ goes to 0

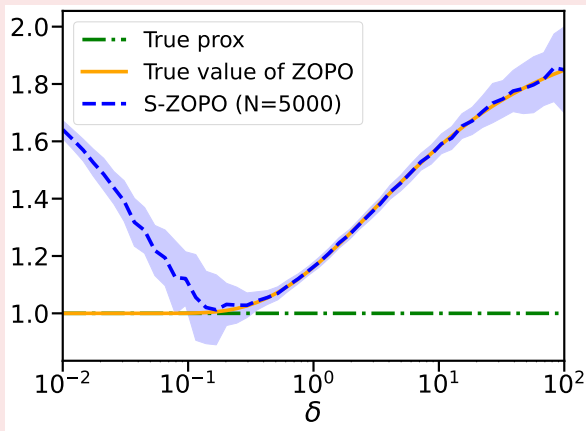
→ Treat $\widehat{\text{zprox}}$ as an inexact prox

The low-temperature regime

What if δ is too small?

The low-temperature regime

What if δ is too small? $z_{\text{prox}}^{\delta}_{\lambda, f}$ becomes harder to estimate!



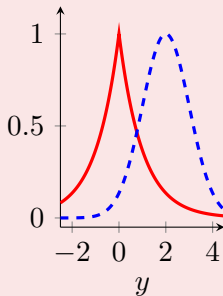
The low-temperature regime

What if δ is too small? $\text{zprox}_{\lambda, f}^{\delta}$ becomes harder to estimate!

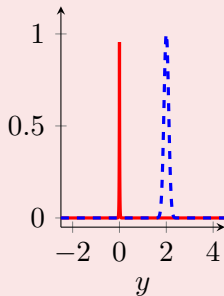
Key quantity

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

(a) Large δ



(b) Small δ



The low-temperature regime

What if δ is too small?

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

The low-temperature regime

What if δ is too small?

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

The low-temperature regime

What if δ is too small?

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

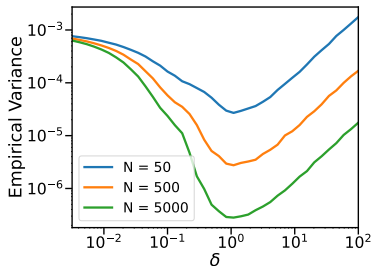
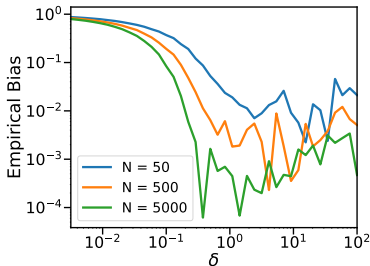
→ Acts as a **random search operator**!

The low-temperature regime

What if δ is too small?

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

→ Acts as a **random search operator**! **Problematic bias and variance!**



Our work

Let's study the high-temperature regime!

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

The high-temperature regime

What happens for a fixed δ ?

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

The high-temperature regime

What happens for a fixed δ ?

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

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

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

→ it is an **MMSE estimator**

The high-temperature regime

What happens for a fixed δ ?

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

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

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

→ it is an **MMSE estimator**

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

Zeroth Order Proximal Point Algorithm

ZOPO is a PO!

- the underlying potential is **explicit**:

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

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

Zeroth Order Proximal Point Algorithm

ZOPO is a PO!

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

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

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

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

Sampled Zeroth Order Proximal Point Algorithm

S-ZOPPA: for a sequence of scheduled batch sizes $(N_k)_k$,

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

Theorem [Naldi, L., Molinari, Villa '26] Suppose that $(x^k)_k$ is bounded a.s. and minimal assumptions on f .

- a) **Convex case:** if $f^{\lambda, \delta}$ is convex and $\sum_k N_k^{-\frac{1}{2}} < \infty$, then $x^k \rightarrow \bar{x} \in \arg \min f^{\lambda, \delta}$ a.s.
- b) **Nonconvex case:** if $\sum_k \frac{k^{2\alpha}}{N_k} < \infty$ for some $\alpha > 0$, then $x^k \rightarrow \bar{x}$ a.s. such that $\nabla f^{\lambda, \delta}(\bar{x}) = 0$.

Hidden ingredients: Kurdyka-Lojasiewicz property of $f^{\lambda, \delta}$, firm nonexpansiveness of zprox in the convex case...

On the viscous Moreau envelope

What can be said about $f^{\lambda,\delta}$?

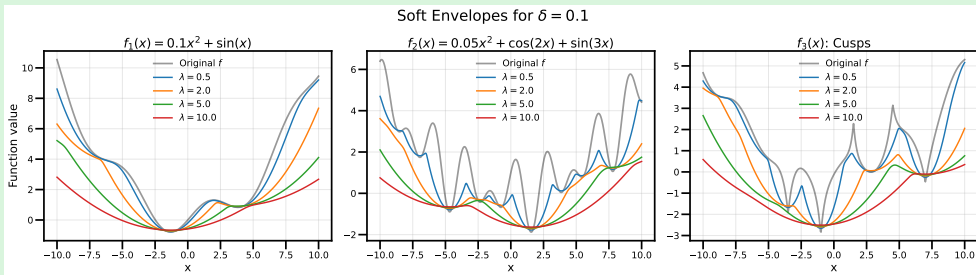
- **Characterize its critical points:** if f is L -smooth and $\lambda < 1/L$ (see also [Lauga, Vaiter '26]),

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

- **Effect of the parameter λ :** under a dissipativity condition, **for a sufficiently large λ , $f^{\lambda,\delta}$ is locally convex!**

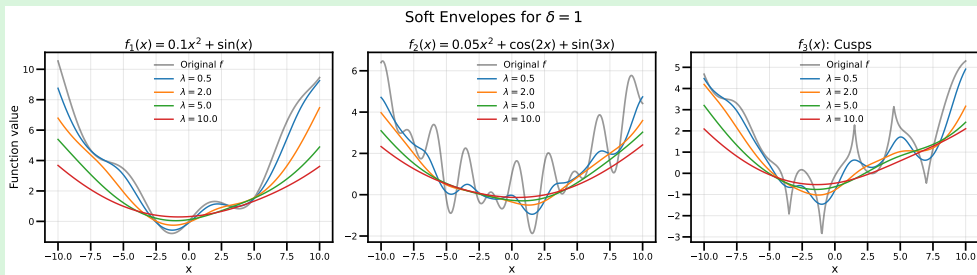
On the viscous Moreau envelope

Convexification effect: increasing λ convexifies $f^{\lambda,\delta}$



On the viscous Moreau envelope

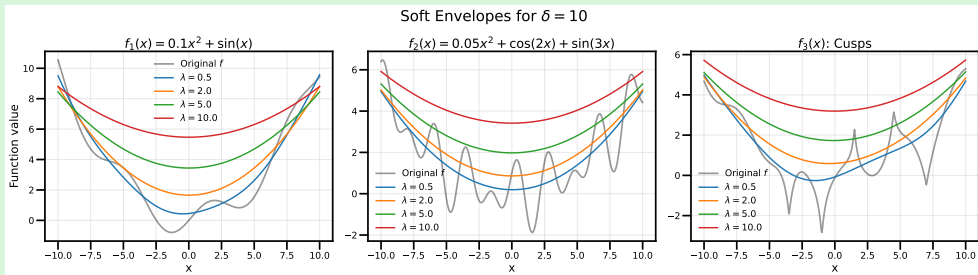
Convexification effect: increasing λ convexifies $f^{\lambda,\delta}$



Even more for larger δ !

On the viscous Moreau envelope

Convexification effect: increasing λ convexifies $f^{\lambda,\delta}$



Even more for larger δ !

Conclusion

Take-away message:

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

Perspectives:

- Leverage theory to develop **efficient ZOPO algorithms** (Graduated optimization, adaptive parameters, better sampling)
- Extend this framework to **particle-based methods** (CBO, PSO) featuring similar objects [Bianchi, Dragomir, Priser '25, Fornasier, Klock, Riedl '24]

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.
- Bianchi, Dragomir, Priser, *Consensus-based optimization beyond finite-time analysis*, 2025.
- Fornasier, Klock, Riedl, *Consensus-based optimization methods converge globally*, SIAM Journal on Optimization, 2024.

Effective Sample Size

What is the ESS? Coming from **Importance Sampling**:

Estimate $\mathbb{E}_{x \sim p}[x]$ by sampling from q :

$$\mathbb{E}_{x \sim p}[x] = \mathbb{E}_{z \sim q} \left[z \frac{p(z)}{q(z)} \right] \approx \frac{1}{N} \sum_{i=1}^N z_i \underbrace{\frac{p(z_i)}{q(z_i)}}_{:=\omega(z_i)}, \quad z_i \sim q$$

$$ESS = \frac{\left(\sum_{i=1}^N \omega(z_i) \right)^2}{\sum_{i=1}^N \omega(z_i)^2}$$

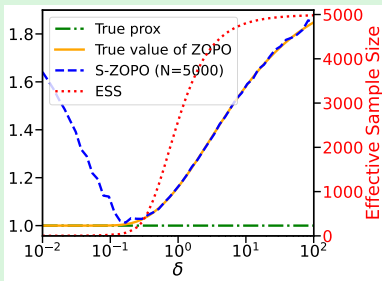
ESS (informal): *For N samples from q , how many samples from p would give a similar variance/statistical efficiency?*

→ ESS equal to 1 = poor surrogate q , ESS equal to N = "perfect" surrogate

Effective Sample Size

ESS for SZOPO: $\widehat{z_{\text{prox}}}_{\lambda, f}^{\delta, N}$ is a **Self-Normalized Importance Sampling** (SNIS) estimator

$$ESS = \frac{\left(\sum_{i=1}^N \exp \left(-\frac{f(y_i)}{\delta} \right) \right)^2}{\sum_{i=1}^N \exp \left(-2\frac{f(y_i)}{\delta} \right)}$$



→ Tracks how bad is the sampled estimate