Erratum: Deep Equilibrium Models for Poisson Imaging Inverse Problems via Mirror Descent

Formalization Blueprint (Lean 4)

November 16, 2026

Abstract

This blueprint accompanies a complete, `sorry`-free Lean 4 formalization (using pure Mathlib v4.27.0) of the erratum to Daniele, Villa, Vaiter, and Calatroni (*SIAM J. Imaging Sci.*, 2026, arXiv:2507.11461) concerning Deep Equilibrium Models for Poisson imaging inverse problems via Mirror Descent. We formalize:

1. the variational calculus primitives and a full proof of the abstract Kurdyka–Łojasiewicz (KŁ) convergence theorem (Theorem A.8 of Attouch, Bolte, and Svaiter);
2. the Poisson/Kullback–Leibler variational setup with zero counts permitted ($y_i \geq 0$), the counterexample (**Example 1**) showing that the classical Burg entropy gradient fails to be Lipschitz continuous on sublevel sets L_0, and the shifted Burg geometry (h_ξ, D_{h_ξ}) yielding the variational equivalence (1) $\iff$ (4) $\iff$ (5);
3. sublevel separation from the boundary singular set and relative smoothness $L_\beta h_\xi - F$ convex on K_β (**Lemma 2**);
4. single-step positivity of denominators, sublevel invariance, Bregman sufficient decrease (6), and relative error bound (7) (**Lemma 3**), followed by constant-step sequence convergence (**Theorem 4**, passing the definable KŁ property HasKL(Ψ) as an explicit hypothesis);
5. monotone backtracking step-size stabilization and global convergence (**Algorithm 1**).

1 Variational Analysis and Abstract KŁ Convergence

Let $\mathcal{E}_n := \mathbb{R}^n$ equipped with the standard Euclidean inner product $\langle \cdot, \cdot \rangle$ and norm $\| \cdot \|$. We work with extended-real functions $f : \mathcal{E}_n \to \mathbb{R} \cup \{-\infty, +\infty\}$.

Definition 1 (Effective domain and properness)**.** For $f : \mathcal{E}_n \to (-\infty, +\infty]$, the *effective domain* is

$$\operatorname{dom}(f) := \{x \in \mathcal{E}_n \mid f(x) \neq +\infty \text{ and } f(x) \neq -\infty\}.$$

The function f is *proper* if $\operatorname{dom}(f) \neq \emptyset$ and $f(x) \neq -\infty$ for all $x \in \mathcal{E}_n$.

Definition 2 (Fréchet subdifferential, limiting subdifferential, and normal cone)**.** Let $f : \mathcal{E}_n \to (-\infty, +\infty]$ and $x \in \operatorname{dom}(f)$. A vector $v \in \mathcal{E}_n$ belongs to the *Fréchet subdifferential* $\hat{\partial} f(x)$ if

$$\liminf_{\substack{y \to x \\ y \neq x}} \frac{f(y) - f(x) - \langle v, y - x \rangle}{\|y - x\|} \geq 0.$$

The *limiting (Mordukhovich) subdifferential* $\partial f(x)$ is the set of $v \in \mathcal{E}_n$ for which there exist sequences $x_k \to x$ and $v_k \in \hat{\partial} f(x_k)$ with $f(x_k) \to f(x)$ and $v_k \to v$. Critical points are defined by $\mathrm{crit}(f) := \{x \in \mathcal{E}_n \mid 0 \in \partial f(x)\}$. For a subset $C \subseteq \mathcal{E}_n$ and $x \in C$, the convex normal cone is

$$N_C(x) := \{v \in \mathcal{E}_n \mid \forall z \in C,\ \langle v, z - x\rangle \le 0\}.$$

Lemma 3 (Closedness of the limiting subdifferential)**.** *If $x_k \to \bar{x}$, $f(x_k) \to f(\bar{x})$ with $\bar{x} \in \mathrm{dom}(f)$, and $\omega_k \in \partial f(x_k)$ satisfies $\omega_k \to \bar{\omega}$, then $\bar{\omega} \in \partial f(\bar{x})$.*

Proof. Each ω_k is approximated arbitrarily closely in both argument, attentive value $f(x_k)$, and subgradient by elements (x_k', ω_k') of the Fréchet subdifferential graph. A diagonal extraction yields $(y_k, v_k) \in \mathrm{gph}(\hat{\partial} f)$ with $y_k \to \bar{x}$, $f(y_k) \to f(\bar{x})$, and $v_k \to \bar{\omega}$, proving $\bar{\omega} \in \partial f(\bar{x})$. □

Lemma 4 (Smooth sum rule on an open neighborhood)**.** *Suppose $U \subseteq \mathcal{E}_n$ is open, $x \in U$, $g : \mathcal{E}_n \to \mathbb{R}$ is C^1 on U, $h : \mathcal{E}_n \to (-\infty, +\infty]$ is proper, and $f = g + h$ on U. Then*

$$\partial f(x) = \{\nabla g(x) + w \mid w \in \partial h(x)\}.$$

Proof. First prove the local Fréchet subdifferential sum rule $\hat{\partial} f(y) = \nabla g(y) + \hat{\partial} h(y)$ for every $y \in U$ via the first-order Taylor remainder bound for g at y. Because g and ∇g are continuous on U, attentive convergence $y_k \to x$ with $f(y_k) \to f(x)$ is equivalent to $h(y_k) \to h(x)$, and $\nabla g(y_k) + v_k \to \bar{\omega}$ corresponds bijectively to $v_k \to \bar{\omega} - \nabla g(x)$. □

Lemma 5 (Limiting subdifferential of closed convex box indicator)**.** *If $C \subseteq \mathcal{E}_n$ is closed and convex, then for all $x \in \mathcal{E}_n$, $\partial \iota_C(x) = N_C(x)$.*

Proof. Convexity implies $\hat{\partial} \iota_C(y) = N_C(y)$ for all $y \in C$, and closedness of C together with the linear-inequality definition of N_C makes the graph of N_C closed. □

Definition 6 (Desingularizing functions and the KŁ property)**.** For $\eta > 0$, a function $\varphi : \mathbb{R} \to \mathbb{R}$ is a *desingularizing function* in Φ_η if $\varphi(0) = 0$, φ is continuous on $[0, \eta)$, C^1 and concave on $(0, \eta)$, and satisfies $\varphi'(s) > 0$ on $(0, \eta)$.

Definition 7 (Kurdyka–Łojasiewicz property)**.** A proper extended-real function f satisfies the *KŁ property at $\bar{x} \in \mathrm{dom}(\partial f)$* if there exist $\eta > 0$, a neighborhood U of $\bar{x}$, and $\varphi \in \Phi_\eta$ such that for every $x \in U$ with $f(\bar{x}) < f(x) < f(\bar{x}) + \eta$,

$$\varphi'((f(x) - f(\bar{x}))_{\mathbb{R}}) \cdot \mathrm{dist}(0, \partial f(x)) \ge 1.$$

We say f satisfies the KŁ property (HasKL(f)) if it has the KŁ property at every $\bar{x} \in \mathrm{dom}(\partial f)$.

Definition 8 (Omega-limit set)**.** The ω-limit set $\omega(x^0)$ of a sequence $(x_k)_{k \in \mathbb{N}}$ in $\mathcal{E}_n$ is the set of all cluster points $\bar{x}$ for which there exists a strictly increasing extraction $\psi : \mathbb{N} \to \mathbb{N}$ with $x_{\psi(j)} \to \bar{x}$.

Lemma 9 (Compactness and attraction of the ω-limit set)**.** *If $(x_k)_{k \in \mathbb{N}}$ is bounded, then $\omega(x^0)$ is nonempty, closed, and compact, and $\mathrm{dist}(x_k, \omega(x^0)) \to 0$ as $k \to \infty$.*

Proof. Bolzano–Weierstrass in finite-dimensional Euclidean space yields nonemptiness and sequential compactness inside the closure of the range. Closedness follows from a diagonal subsequence argument, and if $\mathrm{dist}(x_k, \omega(x^0))$ failed to tend to zero, an ε-separated subsequence would produce a cluster point outside $\omega(x^0)$. □

Theorem 10 (Uniformized KŁ property)**.** *Let $\Omega \subseteq \mathcal{E}_n$ be nonempty and compact, and suppose f is constant on Ω with finite value $f_\star$ and satisfies the KŁ property at every $z \in \Omega$. Then there exist $\varepsilon > 0$, $\eta > 0$, and a common desingularizing function $\varphi \in \Phi_\eta$ such that for all $x \in \mathcal{E}_n$ with* $\operatorname{dist}(x, \Omega) < \varepsilon$ *and $f_\star < f(x) < f_\star + \eta$,*

$$\varphi'((f(x) - f_\star)_{\mathbb{R}}) \cdot \operatorname{dist}(0, \partial f(x)) \geq 1.$$

Proof. Cover compact Ω by finitely many balls $B(z_i, \delta_i/2)$ extracted from the local KŁ neighborhoods $(U_{z_i}, \eta_{z_i}, \varphi_{z_i})$, pick $\varepsilon > 0$ via the Lebesgue number of the open cover, set $\eta := \min_i \eta_{z_i}$, and form the finite sum $\varphi(s) := \sum_{i \in S} \varphi_{z_i}(s)$. Concavity and positivity of derivatives are preserved under finite sums, and $\varphi'(s) \geq \varphi'_{z_i}(s)$ gives the uniform KŁ inequality without requiring Lebesgue integration. □

Theorem 11 (Abstract KŁ convergence — Theorem A.8 of [3])**.** *Let $f : \mathcal{E}_n \to (-\infty, +\infty]$ be proper, lower semicontinuous, bounded below, and satisfy the KŁ property. Let $(x_k)_{k \in \mathbb{N}}$ be a bounded sequence starting at $x_0 \in \operatorname{dom}(f)$ satisfying for constants $c_1, b > 0$:*

(H1) $f(x_{k+1}) + c_1 \|x_{k+1} - x_k\|^2 \leq f(x_k)$ for all $k \in \mathbb{N}$;

(H2) there exists $\omega_{k+1} \in \partial f(x_{k+1})$ with $\|\omega_{k+1}\| \leq b\|x_{k+1} - x_k\|$ for all $k \in \mathbb{N}$;

(H3) whenever $x_{\psi(j)} \to \bar{x}$, we have $f(x_{\psi(j)}) \to f(\bar{x})$.

Then $\sum_{k=0}^{\infty} \|x_{k+1} - x_k\| < +\infty$, and (x_k) converges to a critical point $x_\infty \in \operatorname{crit}(f)$.

Proof. By (H1) and lower boundedness, $f(x_k) \downarrow f_\star$ and $\|x_{k+1} - x_k\| \to 0$. Using (H2), (H3), and Lemma 3, every $\bar{x} \in \omega(x^0)$ satisfies $\bar{x} \in \operatorname{crit}(f)$ and $f(\bar{x}) = f_\star$. If $f(x_{k_0}) = f_\star$ at some step, monotonicity forces $x_k = x_{k_0}$ for all $k \geq k_0$. Otherwise, applying Theorem 10 on $\Omega = \omega(x^0)$ and concavity of φ yields, for $k \geq K$,

$$2\|x_{k+1} - x_k\| \leq \|x_k - x_{k-1}\| + \frac{b}{c_1}(\varphi(f(x_k) - f_\star) - \varphi(f(x_{k+1}) - f_\star)).$$

Summing gives $\sum \|x_{k+1} - x_k\| < +\infty$, hence Cauchyness and convergence of (x_k) to a limit $x_\infty \in \omega(x^0) \subseteq \operatorname{crit}(f)$. □

2 Poisson Problem Setup, Counterexample, and Shifted Burg Geometry

Definition 12 (Box constraint, KL divergence, and problem data)**.** Let $a > 0$, $\xi > 0$, $y \in \mathbb{R}^m_{\geq 0}$ (zero counts allowed), $A \in \mathbb{R}^{m \times n}_{\geq 0}$ with every row nonzero, and $R_\theta : \mathcal{E}_n \to \mathbb{R}$ of class C^2 on an open neighborhood $U_R \supseteq [0, a]^n$. Set $C := [0, a]^n$,

$$\operatorname{KL}(y, Ax) := \sum_{i=1}^{m} d_{\operatorname{KL}}(y_i, (Ax)_i), \qquad F(x) := \sum_{i=1}^{m} (y_i \log \tfrac{y_i}{(Ax)_i} + (Ax)_i - y_i) + R_\theta(x),$$

with $0 \log(0/u) := 0$ (even for $u = 0$), and $\Psi(x) := \operatorname{KL}(y, Ax) + R_\theta(x) + \iota_{[0,a]^n}(x)$. Define the open positive-observation set $U_+ := \{x \in \mathcal{E}_n \mid \forall i,\ y_i > 0 \implies (Ax)_i > 0\}$ and the boundary singular set

$$\mathcal{S} := \{x \in [0, a]^n \mid \exists i \in \{1, \dots, m\},\ y_i > 0 \text{ and } (Ax)_i = 0\}.$$

Lemma 13 (Effective domain of Ψ and singular boundary)**.** *We have* $\mathrm{dom}(\Psi) = [0,a]^n \cap U_+$, $\overline{\mathrm{dom}(\Psi)} = [0,a]^n$, *and* $\overline{\mathrm{dom}(\Psi)} \setminus \mathrm{dom}(\Psi) = \mathcal{S}$.

Proof. Membership $(Ax)_i > 0$ for every i with $y_i > 0$ characterizes finiteness of $\mathrm{KL}(y, Ax)$, while $\iota_{[0,a]^n}(x) < +\infty$ characterizes $x \in [0,a]^n$. Approximating any $x \in [0,a]^n$ along convex combinations $(1-t)x + t(a, \dots, a)$ stays in $[0,a]^n \cap U_+$ for $t \in (0,1]$, proving $\overline{\mathrm{dom}(\Psi)} = [0,a]^n$ and $\overline{\mathrm{dom}(\Psi)} \setminus \mathrm{dom}(\Psi) = \mathcal{S}$. □

Definition 14 (Classical and shifted Burg entropies and Bregman divergence)**.** The classical Burg entropy on $\mathcal{E}_n$ is $h(x) := -\sum_{j=1}^n \log x_j$ on $\mathbb{R}^n_{>0}$ (and $+\infty$ otherwise), with $(\nabla h(x))_j = -1/x_j$. For shift $\xi > 0$, the shifted Burg entropy and Bregman divergence are

$$h_\xi(x) := \sum_{j=1}^n (-\log(x_j + \xi)), \qquad (\nabla h_\xi(x))_j = -\frac{1}{x_j + \xi},$$

$$D_{h_\xi}(u, x) := h_\xi(u) - h_\xi(x) - \langle \nabla h_\xi(x), u - x \rangle.$$

Example 15 (Counterexample to Lipschitz continuity of ∇h on L_0 — Example 1)**.** With $A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, $R_\theta \equiv 0$, $a \geq 1$, $y \in \mathbb{R}^2_{>0}$, and $x^0 = (0,1)$, we have $\Psi(x^0) < +\infty$ and every $x_t := (t,1)$ for $t \in (0,1]$ lies in $L_0(x^0) \cap (0,a]^n$ with $t \mapsto \Psi(x_t)$ bounded above on $(0,1]$. Yet $\|\nabla h(x_t) - \nabla h(x_1)\| = |1 - 1/t| \to +\infty$ while $\|x_t - x_1\| = 1 - t \leq 1$, so ∇h is not Lipschitz continuous on $\mathrm{dom}(h) \cap L_0(x^0)$.

Proof. Direct evaluation gives $Ax_t = (t+1, 1)$ componentwise ≥ 1, keeping $\mathrm{KL}(y, Ax_t)$ continuous on $[0,1]$, hence bounded by its maximum. Meanwhile $\nabla h(x_t)_0 - \nabla h(x_1)_0 = 1 - 1/t$ grows arbitrarily large as $t \downarrow 0$ whereas $\|x_t - x_1\| \leq 1$. □

Lemma 16 (Strong convexity and Lipschitz gradient of h_ξ on $[0,a]^n$)**.** *For all* $u, x \in [0,a]^n$,

$$D_{h_\xi}(u, x) \geq \frac{1}{2(a+\xi)^2} \|u - x\|^2, \qquad \|\nabla h_\xi(u) - \nabla h_\xi(x)\| \leq \frac{1}{\xi^2} \|u - x\|.$$

Proof. By the 1D integral/mean-value identity for $s \mapsto -\log(s + \xi)$ on $[0,a]$, each coordinate Bregman divergence satisfies $-\log(\frac{u_j+\xi}{x_j+\xi}) + \frac{u_j - x_j}{x_j+\xi} \geq \frac{(u_j - x_j)^2}{2(a+\xi)^2}$, and $|-(u_j+\xi)^{-1} + (x_j+\xi)^{-1}| \leq \frac{|u_j - x_j|}{\xi^2}$. □

Definition 17 (Mirror Descent subproblem and closed-form update T_τ)**.** For step size $\tau > 0$ and $x \in [0,a]^n$, the shifted-Burg Mirror Descent subproblem objective is

$$\Phi_{x,\tau}(z) := \iota_{[0,a]^n}(z) + \langle z - x, \nabla F(x) \rangle + \frac{1}{\tau} D_{h_\xi}(z, x).$$

We define the operator T_τ coordinatewise by

$$(T_\tau(x))_j := \begin{cases} \Pi_{[0,a]}\left(\dfrac{x_j + \xi}{1 + \tau(x_j + \xi)\partial_j F(x)} - \xi \right) & \text{if } 1 + \tau(x_j + \xi)\partial_j F(x) > 0, \\ a & \text{otherwise.} \end{cases}$$

Theorem 18 (Variational equivalence (1) $\iff$ (4) $\iff$ (5) and continuity of T_τ)**.** *For any* $\tau > 0$ *and* $x \in [0,a]^n$ *satisfying* $1 + \tau(x_j + \xi)\partial_j F(x) > 0$ *for all* j*:*

1. *$T_\tau(x)$ satisfies the optimality inclusion (5):*

$$-\nabla F(x) - \frac{1}{\tau}(\nabla h_\xi(T_\tau(x)) - \nabla h_\xi(x)) \in \partial \iota_{[0,a]^n}(T_\tau(x)) = N_{[0,a]^n}(T_\tau(x));$$

2. *$T_\tau(x)$ is the unique global minimizer (1) of $\Phi_{x,\tau}$ over $\mathcal{E}_n$, and conversely any global minimizer of $\Phi_{x,\tau}$ coincides with $T_\tau(x)$ (4).*

Furthermore, for every $\tau \in \mathbb{R}$, T_τ is continuous on $[0,a]^n \cap U_+$.

Proof. The coordinate scalar inclusion for $N_{[0,a]}(y_j)$ follows by inspecting the three cases $(x_j + \xi)/d_j - \xi \leq 0, \geq a$, and $\in (0,a)$. The three-point Bregman identity together with Lemma 16 makes $\Phi_{x,\tau}(z) - \Phi_{x,\tau}(T_\tau(x)) > 0$ for all $z \neq T_\tau(x)$. Clamping the denominator below by $\delta = \frac{\xi}{2(a+\xi)}$ agrees identically with T_τ on $[0,a]^n \cap U_+$, establishing continuity even when denominators approach or cross zero. □

3 Sublevel Separation and Relative Smoothness — Lemma 2

Definition 19 (Sublevel sets L_0, K_β, and thresholds β_0, β_1)**.** Given an initial point $x^0 \in [0,a]^n$ with $\Psi(x^0) < +\infty$, define

$$L_0(x^0) := \{x \in [0,a]^n \mid \Psi(x) \leq \Psi(x^0)\}, \qquad \beta_0 := \Psi(x^0)_{\mathbb{R}} - \inf_{z \in [0,a]^n} R_\theta(z), \qquad \beta_1 := \beta_0 + 1,$$

and for $\beta \in \mathbb{R}$,

$$K_\beta := \{x \in [0,a]^n \mid \mathrm{KL}(y, Ax) \leq \beta\}.$$

Lemma 20 (Properness, lower semicontinuity, and lower boundedness of Ψ)**.** *Ψ is proper, lower semicontinuous on $\mathcal{E}_n$, and bounded below on $\mathcal{E}_n$ by $\inf_{z \in [0,a]^n} R_\theta(z)$.*

Proof. Nonemptiness of $\mathrm{dom}(\Psi)$ holds at $(a, \dots, a)$, each scalar KL term is nonnegative and lower semicontinuous on $\mathbb{R}$ (including $y_i = 0$ and $u_i \leq 0$), and R_θ is continuous on compact $[0,a]^n$. □

Lemma 21 (Compactness, convexity, and inclusion $L_0 \subseteq K_{\beta_0}$)**.** *For every $\beta \in \mathbb{R}$, K_β is convex, closed, and compact, and $L_0(x^0) \subseteq K_{\beta_0}$.*

Proof. Convexity follows from concavity of log on $(0,\infty)$ applied to $(Ax)_i$. Lower semicontinuity of $\mathrm{KL}(y, A\cdot)$ shows $K_\beta = [0,a]^n \cap \mathrm{KL}(y, A\cdot)^{-1}((-\infty, \beta])$ is closed inside compact $[0,a]^n$. Comparing $\mathrm{KL}(y, Ax) \leq \Psi(x)_{\mathbb{R}} - R_\theta(x)$ gives $L_0(x^0) \subseteq K_{\beta_0}$. □

Lemma 22 (Positive lower bound and separation from $\mathcal{S}$)**.** *For every $\beta \in \mathbb{R}$, there exists $c_\beta > 0$ such that $(Ax)_i \geq c_\beta$ for all $x \in K_\beta$ and every i with $y_i > 0$. Consequently, $K_\beta \subseteq U_+ \cap U_R$ and $K_\beta \cap \mathcal{S} = \emptyset$.*

Proof. Nonnegativity of each scalar KL term yields $d_{\mathrm{KL}}(y_i, (Ax)_i) \leq \max(\beta, 0)$ on K_β, forcing $(Ax)_i \geq y_i \exp(-(\max(\beta,0) + y_i)/y_i) > 0$ whenever $y_i > 0$. □

Definition 23 (Smoothness constants M_β and L_β)**.** For $\beta \in \mathbb{R}$, define

$$M_\beta := 1 + \sup_{x \in K_\beta} \|D(\nabla F)(x)\|_{\mathrm{op}}, \qquad L_\beta := M_\beta (a + \xi)^2.$$

Lemma 24 (Lipschitz continuity of ∇F on K_β)**.** *F is C^2 on the open set $U_+ \cap U_R \supseteq K_\beta$, and ∇F is M_β-Lipschitz on the compact convex set K_β:*

$$\forall u, x \in K_\beta, \qquad \|\nabla F(u) - \nabla F(x)\| \le M_\beta \|u - x\|.$$

Proof. Compactness of K_β and continuity of $D(\nabla F)$ on $U_+ \cap U_R$ bound the operator norm uniformly by M_β on K_β; convexity of K_β allows the mean value theorem along line segments inside K_β. □

Theorem 25 (Relative smoothness — Lemma 2)**.** *For every $\beta \in \mathbb{R}$, the function $L_\beta h_\xi - F$ is convex on K_β, and hence for all $x, u \in K_\beta$,*

$$F(u) \le F(x) + \langle u - x, \nabla F(x)\rangle + L_\beta D_{h_\xi}(u, x).$$

Proof. Along any segment $z(t) = (1-t)x + tu \subseteq K_\beta$, the derivative $\varphi'(t) = \langle L_\beta \nabla h_\xi(z(t)) - \nabla F(z(t)), u - x\rangle$ satisfies $\varphi'(t_2) - \varphi'(t_1) \ge (L_\beta/(a+\xi)^2 - M_\beta)(t_2 - t_1)\|u - x\|^2 = 0$ for $0 \le t_1 \le t_2 \le 1$, hence is monotone. Applying the mean value theorem along $t \in [0, 1]$ yields the Bregman descent inequality. □

4 Lemma 3 and Constant-Step Convergence — Theorem 4

Definition 26 (Step-size threshold $\bar\tau$ and constant-step iterates)**.** Set $G := \sup_{x \in K_{\beta_0}} \|\nabla F(x)\|_\infty \ge 0$ and

$$\bar\tau := \min\left\{ \frac{1}{(a+\xi)G}, \frac{1}{L_{\beta_1}} \right\} > 0$$

(taking $\bar\tau := 1/L_{\beta_1}$ if $G = 0$).

Theorem 27 (Single-step well-definedness, descent (6), and relative error (7) — Lemma 3)**.** *Let $x^0 \in [0, a]^n$ with $\Psi(x^0) < +\infty$. For every $x \in L_0(x^0)$ and every $\tau \in (0, \bar\tau)$:*

1. *$1 + \tau(x_j + \xi)\partial_j F(x) > 0$ for every coordinate $j \in \{1, \dots, n\}$;*
2. *$T_\tau(x) \in K_{\beta_1} \cap L_0(x^0)$;*
3. *Bregman sufficient decrease (6) holds:*

$$\Psi(x) - \Psi(T_\tau(x)) \ge \left(\frac{1}{\tau} - L_{\beta_1}\right) D_{h_\xi}(T_\tau(x), x);$$

4. *setting $\omega := \nabla F(T_\tau(x)) - \nabla F(x) + \frac{1}{\tau}(\nabla h_\xi(x) - \nabla h_\xi(T_\tau(x)))$, we have $\omega \in \partial\Psi(T_\tau(x))$ and the relative error bound (7):*

$$\|\omega\| \le \left(M_{\beta_1} + \frac{1}{\tau\xi^2}\right) \|T_\tau(x) - x\|.$$

Proof. Using the continuous map $t \mapsto T_t(x)$ on $[0, \tau]$, define $U := \{t \in [0, \tau] \mid T_t(x) \in K_{\beta_0}\}$ (closed, containing 0) and $V := \{t \in \mathbb{R} \mid T_t(x) \in U_+ \text{ and } \mathrm{KL}(y, AT_t(x)) < \beta_1\}$ (open). Any $t \in [0, \tau] \cap V$ has $T_t(x) \in K_{\beta_1}$ and $1 + t(x_j + \xi)\partial_j F(x) > 0$, so Theorem 25 and Theorem 18 give $\Psi(T_t(x)) \le \Psi(x) \le \Psi(x^0)$, putting $t \in U$. By connectedness of $[0, \tau]$, $\tau \in U$, yielding $T_\tau(x) \in K_{\beta_0} \subseteq K_{\beta_1}$, (6), and $T_\tau(x) \in L_0(x^0)$. Combining Theorem 18(1) with Lemma 4 and Lipschitz bounds for ∇F and ∇h_ξ gives (7). □

Definition 28 (Constant-step Mirror Descent sequence)**.** For fixed $\tau > 0$ and $x^0 \in [0,a]^n$, define $x^0_{\mathrm{MD}} := x^0$ and $x^{k+1} := T_\tau(x^k)$ for $k \in \mathbb{N}$.

Theorem 29 (Global convergence for constant step size — Theorem 4)**.** *Let $x^0 \in [0,a]^n$ with $\Psi(x^0) < +\infty$, let $\tau \in (0, \bar{\tau})$, and assume Ψ satisfies the KŁ property (*HasKL(Ψ)*). Then the sequence $(x^k)_{k\in\mathbb{N}}$ satisfies*

$$\sum_{k=0}^{\infty} \|x^{k+1} - x^k\| < +\infty$$

and converges to a critical point $x^\infty \in \mathrm{crit}(\Psi) \cap L_0(x^0)$ which is a fixed point of T_τ, i.e., $x^\infty = T_\tau(x^\infty)$.

Proof. By induction via Theorem 27, every x^k stays in the bounded closed set $L_0(x^0) \subseteq K_{\beta_0}$. Combining (6) with Lemma 16 verifies (H1) with $c_1 = \frac{1/\tau - L_{\beta_1}}{2(a+\xi)^2} > 0$, while (7) gives (H2) with $b = M_{\beta_1} + \frac{1}{\tau\xi^2} > 0$, and continuity of F on $U_+ \cap U_R \supseteq K_{\beta_0}$ verifies (H3). Theorem 11 yields summability and convergence $x^k \to x^\infty \in \mathrm{crit}(\Psi) \cap L_0(x^0)$, and continuity of T_τ on $[0,a]^n \cap U_+$ passes to the limit $x^\infty = T_\tau(x^\infty)$. □

5 Monotone Backtracking Mirror Descent — Algorithm 1

Definition 30 (Backtracking threshold $\hat{\tau}$, acceptance criterion, and Algorithm 1 sequence)**.** For $\gamma \in (0,1)$, $\eta \in (0,1)$, $\tau_0 > 0$, and $x^0 \in L_0(x^0)$, set

$$\hat{\tau} := \min\left\{ \frac{1}{(a+\xi)G},\ \frac{1-\gamma}{L_{\beta_1}} \right\} \in (0, \bar{\tau}].$$

A step size $\tau > 0$ is *accepted* at $x \in L_0(x^0)$ if $1 + \tau(x_j + \xi)\partial_j F(x) > 0$ for all j, $T_\tau(x) \in \mathrm{dom}(\Psi)$, and

$$\Psi(x) - \Psi(T_\tau(x)) \geq \frac{\gamma}{\tau} D_{h_\xi}(T_\tau(x), x).$$

Starting from $(x^0, \tau_{-1}) := (x^0, \tau_0)$, Algorithm 1 sets $\tau_k := \eta^{m_k}\tau_{k-1}$ where $m_k \in \mathbb{N}$ is the smallest nonnegative integer for which τ_k is accepted at x^k, and $x^{k+1} := T_{\tau_k}(x^k)$.

Lemma 31 (Acceptance below $\hat{\tau}$ and finite inner termination)**.** *Every $\tau \in (0, \hat{\tau})$ is accepted at any $x \in L_0(x^0)$. Consequently, for any $\tau_{\mathrm{prev}} > 0$, there exists $m \in \mathbb{N}$ accepting $\eta^m\tau_{\mathrm{prev}}$, and the resulting monotone step sizes satisfy $\min\{\tau_0, \eta\hat{\tau}\} \leq \tau_k \leq \tau_0$ for all $k \in \mathbb{N}$.*

Proof. For $\tau \in (0, \hat{\tau}) \subseteq (0, \bar{\tau})$, $\tau L_{\beta_1} \leq 1 - \gamma$ gives $1/\tau - L_{\beta_1} \geq \gamma/\tau$, so Theorem 27 yields acceptance. Since $\eta^m\tau_{\mathrm{prev}} \to 0$, backtracking terminates finitely, and if $m_k \geq 1$ then failure at $m_k - 1$ implies $\eta^{m_k-1}\tau_{k-1} \geq \hat{\tau}$, hence $\tau_k \geq \eta\hat{\tau}$. □

Theorem 32 (Finite stabilization of the backtracking step size)**.** *The step-size sequence $(\tau_k)_{k\in\mathbb{N}}$ is monotonically nonincreasing and stabilizes in finitely many outer iterations: there exist $k_{\mathrm{stab}} \in \mathbb{N}$ and $\tau_\infty > 0$ such that $\tau_k = \tau_\infty$ for all $k \geq k_{\mathrm{stab}}$.*

Proof. Every strict decrease multiplies τ_k by at most $\eta < 1$, whereas $\tau_k \geq \tau_{\min} := \min\{\tau_0, \eta\hat{\tau}\} > 0$ uniformly by Lemma 31. Thus at most $N_{\max}$ strict drops can occur, forcing eventual constancy. □

Theorem 33 (Global convergence of Algorithm 1)**.** *Assume Ψ satisfies* HasKL(Ψ)*. Then the iterates $(x^k)_{k\in\mathbb{N}}$ and step sizes $(\tau_k)_{k\in\mathbb{N}}$ generated by Algorithm 1 satisfy:*

1. $\tau_k \downarrow \tau_\infty > 0$ *in finitely many steps, with* $\min\{\tau_0, \eta\hat{\tau}\} \le \tau_k \le \tau_0$;
2. $\sum_{k=0}^{\infty} \|x^{k+1} - x^k\| < +\infty$;
3. $x^k \to x^\infty \in \mathrm{crit}(\Psi) \cap L_0(x^0)$, *and* $x^\infty = T_{\tau_\infty}(x^\infty)$.

Proof. Acceptance gives $\Psi(x^{k+1}) \le \Psi(x^k) \le \Psi(x^0)$, keeping $x^k \in L_0(x^0) \subseteq K_{\beta_0} \subseteq K_{\beta_1}$ for all k. Using $\tau_k \le \tau_0$ and Lemma 16 yields (H1) with uniform constant $c_1 = \frac{\gamma/\tau_0}{2(a+\xi)^2} > 0$. Using the positivity of the accepted denominators together with Theorem 18, Lemma 4, and $\tau_k \ge \tau_{\min} > 0$ yields (H2) with $b = M_{\beta_1} + \frac{1}{\tau_{\min}\xi^2} > 0$. Verification of (H3) and Theorem 11 give finite length and $x^k \to x^\infty \in \mathrm{crit}(\Psi) \cap L_0(x^0)$, while joint continuity of $(\tau, x) \mapsto T_\tau(x)$ and $\tau_k \to \tau_\infty$ yield $x^\infty = T_{\tau_\infty}(x^\infty)$. □