Pseudo-Riemannian optimization for Boolean satisfiability

Mastane Achab*
DeepGambit

Abstract

We generalize the convexity inequality "function above tangent" to a class of functions that are convex in a pseudo-Riemannian sense. We propose a new descent algorithm for such functions and provide a regret analysis. Finally, we explain how to apply that method to design a 3-SAT solver based on continuous relaxation.

Notations

- Interior of simplex: let

$$\mathcal{S} = \left\{ q = (q_0, \dots, q_K) \in (0,1)^{K+1}, \sum_{k=0}^{K} q_k = 1 \right\}.$$

- Shannon entropy: for $q \in \mathcal{S}$,

$$H(q) = -\sum_{k=0}^{K} q_k \log(q_k).$$

- Kullback-Leibler divergence: for $q, r \in \mathcal{S}$,

$$D_{\mathrm{KL}}(q\|r) = \sum_{k=0}^{K} q_k \log\left(\frac{q_k}{r_k}\right) = -H(q) \underbrace{-\sum_{k} q_k \log(r_k)}_{\text{cross-entropy}}.$$

- Softargmax: let the softargmax $\sigma = (\sigma_0, \dots, \sigma_K)$ be defined for all $\xi = (\xi_1, \dots, \xi_K) \in \mathbb{R}^K$ by

$$\sigma_0(\xi) = \frac{1}{1 + \sum_{j=1}^{K} e^{\xi_j}} \quad \text{and} \quad \sigma_k(\xi) = \frac{e^{\xi_k}}{1 + \sum_{j=1}^{K} e^{\xi_j}} \quad \text{for } k \in \{1, \dots, K\}.$$

*mastane@deepgambit.com

- Fisher information matrix: the Fisher information matrix $\mathbf{I}_\sigma$ of the family of distributions $\{\sigma(\xi) : \xi \in \mathbb{R}^K\}$ is the $K \times K$ positive definite matrix given by

$$\boxed{\mathbf{I}_\sigma(\xi) = \mathrm{Diag}(\widetilde{\sigma}(\xi)) - \widetilde{\sigma}(\xi)\widetilde{\sigma}(\xi)^\intercal} \quad \text{with} \quad \boxed{\widetilde{\sigma} = (\sigma_1, \ldots, \sigma_K)}.$$

- Mirror map: let the Bregman generator function Ψ be defined for $(z, \xi) \in \mathbb{R}^n \times \mathbb{R}^K$ as

$$\Psi(z, \xi) = \frac{1}{2}\|z\|^2 + \underbrace{\log\left(1 + \sum_{k=1}^{K} e^{\xi_k}\right)}_{\text{softmax}}$$

and the mirror map

$$\boxed{\nabla\Psi(z, \xi) = \begin{pmatrix} z \\ \widetilde{\sigma}(\xi) \end{pmatrix}}.$$

1 Pseudo-Riemannian convexity

1.1 Sum of log-concave functions

Let $n \geqslant 1$ and $K \geqslant 0$. Consider the optimization problem

$$\max_{z \in \mathbb{R}^n} \sum_{k=0}^{K} p_k(z)$$

where each $p_k : \mathbb{R}^n \to (0, +\infty)$ is log-concave (e.g. product of softargmax entries) and continuously differentiable. This is equivalent to

$$\min_{z \in \mathbb{R}^n} -\log\left(\sum_{k=0}^{K} p_k(z)\right).$$

1.2 Belief state

If $K = 0$, then we have $-\log(p_0)$ that is convex. But in general for $K \geqslant 1$, our objective function is not convex. The natural[1] convex relaxation consists in introducing a "belief state" parameter $\xi \in \mathbb{R}^K$ through the following cross-entropy regularized objective:

$$-\log\left(\sum_{k=0}^{K} p_k(z)\right) + D_{\mathrm{KL}}\left(\sigma(\xi) \middle\| \left(\frac{p_k(z)}{\sum_{j=0}^{K} p_j(z)}\right)_k\right) + H(\sigma(\xi))$$
$$= -\sum_{k=0}^{K} \sigma_k(\xi) \log(p_k(z))$$

which is obviously convex with respect to z.

[1] see Achab [2023] for further motivation

1.3 Pseudo-Riemannian gradient

From now on, we assume that $K \geqslant 1$ and we seek to minimize

$$\boxed{F(z,\xi) = \sigma(\xi)^\intercal \ell(z)} \quad \text{where} \quad \ell_k = -\log(p_k).$$

As we already saw, each ℓ_k is (by definition) convex and thus $z \mapsto F(z,\xi)$ is convex too. But in fact, we have a much stronger notion of joint convexity in terms of the pair (z,ξ) living in the pseudo-Riemannian space $(\mathbb{R}^{n+K}, G)$ equipped with the metric tensor

$$\boxed{G(\xi) = \begin{pmatrix} I_n & 0 \\ 0 & -\mathbf{I}_\sigma(\xi) \end{pmatrix}}.$$

Indeed, the pseudo-Riemannian gradient

$$\nabla^G F(z,\xi) = G(\xi)^{-1} \nabla F(z,\xi)$$

is a Bregman monotone operator, as shown in the next theorem[2].

Hence, the function F is convex in a pseudo-Riemannian sense!

Theorem 1 (Monotonicity)**.** *For all z, ξ, z', ξ',*

$$\left\langle \nabla^G F(z,\xi) - \nabla^G F(z',\xi')\,,\, \nabla\Psi(z,\xi) - \nabla\Psi(z',\xi') \right\rangle \geqslant 0.$$

Proof. Denoting $\widetilde{\ell} = (\widetilde{\ell}_1, \ldots, \widetilde{\ell}_K)$ with

$$\widetilde{\ell}_k = \ell_k - \ell_0,$$

and J_ℓ the Jacobian matrix of ℓ, our pseudo-Riemannian gradient can be rewritten as

$$\nabla^G F(z,\xi) = \begin{pmatrix} J_\ell(z)^\intercal \sigma(\xi) \\ -\widetilde{\ell}(z) \end{pmatrix} \quad \text{where} \quad J_\ell(z)^\intercal \sigma(\xi) = \sum_{k=0}^{K} \sigma_k(\xi) \nabla \ell_k(z).$$

By convexity, we have the inequalities

$$\begin{cases} \sigma(\xi)^\intercal \ell(z') - \sigma(\xi)^\intercal \ell(z) \geqslant \sigma(\xi)^\intercal J_\ell(z)(z'-z) \\ \sigma(\xi')^\intercal \ell(z) - \sigma(\xi')^\intercal \ell(z') \geqslant \sigma(\xi')^\intercal J_\ell(z')(z-z') \end{cases}$$

that we sum to obtain

$$\begin{aligned} &\left\langle \begin{pmatrix} z'-z \\ \sigma(\xi') - \sigma(\xi) \end{pmatrix}, \begin{pmatrix} J_\ell(z')^\intercal \sigma(\xi') - J_\ell(z)^\intercal \sigma(\xi) \\ -\ell(z') + \ell(z) \end{pmatrix} \right\rangle \\ &\qquad = \left\langle \begin{pmatrix} z'-z \\ \widetilde{\sigma}(\xi') - \widetilde{\sigma}(\xi) \end{pmatrix}, \begin{pmatrix} J_\ell(z')^\intercal \sigma(\xi') - J_\ell(z)^\intercal \sigma(\xi) \\ -\widetilde{\ell}(z') + \widetilde{\ell}(z) \end{pmatrix} \right\rangle \geqslant 0 \end{aligned}$$

which concludes the proof. □

[2] originally proved in Proposition 5-(i) in Achab [2024]

1.4 Generalized convexity inequality

Definition 2 (Hourglass gradient)**.** Given z, ξ and z', we define the z'-hourglass gradient of F at (z, ξ) as

$$\eth_{z'} F(z,\xi) = \begin{pmatrix} J_\ell(z)^\intercal \sigma(\xi) \\ 2\int_{\tau=0}^1 \widetilde{\ell}(z+\tau(z'-z))d\tau - \widetilde{\ell}(z) \end{pmatrix}.$$

Lemma 3. *For all z, ξ, z', ξ',*

$$F(z',\xi') - F(z,\xi) \geqslant \left\langle \eth_{z'} F(z,\xi)\,,\, \begin{pmatrix} z'-z \\ \widetilde{\sigma}(\xi') - \widetilde{\sigma}(\xi) \end{pmatrix} \right\rangle.$$

Proof. Let us introduce

$$\widetilde{\nabla} F(z,\xi) = \begin{pmatrix} J_\ell(z)^\intercal \sigma(\xi) \\ \widetilde{\ell}(z) \end{pmatrix}$$

which corresponds to the "natural gradient" (see Amari [2016]) of F under the Riemannian metric tensor $\begin{pmatrix} I_n & 0 \\ 0 & \mathbf{I}_\sigma(\xi) \end{pmatrix}$. Let us consider the function[3]

$$f(\tau) = F\left(z+\tau(z'-z), \widetilde{\sigma}^{-1}(\varphi(\tau))\right) \quad \text{with} \quad \varphi(\tau) = \widetilde{\sigma}(\xi) + \tau(\widetilde{\sigma}(\xi') - \widetilde{\sigma}(\xi)).$$

Then,

$$\begin{aligned} F(z',\xi') - F(z,\xi) &= f(1) - f(0) = \int_{\tau=0}^1 f'(\tau)d\tau = \left\langle \widetilde{\nabla} F(z,\xi)\,,\, \begin{pmatrix} z'-z \\ \widetilde{\sigma}(\xi') - \widetilde{\sigma}(\xi) \end{pmatrix} \right\rangle \\ &+ \int_{\tau=0}^1 \frac{1}{\tau} \left\langle \widetilde{\nabla} F\left(z+\tau(z'-z), \widetilde{\sigma}^{-1}(\varphi(\tau))\right) - \widetilde{\nabla} F\left(z,\xi\right)\,,\, \tau \begin{pmatrix} z'-z \\ \widetilde{\sigma}(\xi') - \widetilde{\sigma}(\xi) \end{pmatrix} \right\rangle d\tau \\ &\geqslant \left\langle \widetilde{\nabla} F(z,\xi)\,,\, \begin{pmatrix} z'-z \\ \widetilde{\sigma}(\xi') - \widetilde{\sigma}(\xi) \end{pmatrix} \right\rangle + 2\int_{\tau=0}^1 \left\langle \widetilde{\ell}\left(z+\tau(z'-z)\right) - \widetilde{\ell}(z)\,,\, \widetilde{\sigma}(\xi') - \widetilde{\sigma}(\xi) \right\rangle d\tau, \end{aligned}$$

where the inequality results from applying Theorem 1 inside the integral. □

2 Hourglass Descent

2.1 Method

Given step-sizes $\lambda_t, \beta_t > 0$, we define the Hourglass Descent iteration as,

$$\begin{pmatrix} z^t \\ \xi^t \end{pmatrix} = \begin{pmatrix} z^{t-1} \\ \xi^{t-1} \end{pmatrix} - \begin{pmatrix} \lambda_t I_n & 0 \\ 0 & \beta_t I_K \end{pmatrix} \eth_{z^t} F(z^{t-1}, \xi^{t-1})\,. \tag{HD}$$

[3]Note that $\widetilde{\sigma}$ is a bijection from $\mathbb{R}^K$ to $\left\{q = (q_1, \ldots, q_K) \in (0,1)^K : \sum_{j=1}^K q_j < 1\right\}$ with inverse $\widetilde{\sigma}^{-1}(q) = \left(\log \frac{q_k}{1-\sum_{j=1}^K q_j}\right)_{1 \leqslant k \leqslant K}$.

Though z^t appears on both sides of (HD), we can decompose the update rule explicitly as

$$\begin{cases} z^t = z^{t-1} - \lambda_t J_\ell(z^{t-1})^\intercal \sigma(\xi^{t-1}) \\ \xi^t = \xi^{t-1} - \beta_t \left(2\int_{\tau=0}^1 \widetilde{\ell}(z^{t-1} - \tau\lambda_t J_\ell(z^{t-1})^\intercal \sigma(\xi^{t-1}))d\tau - \widetilde{\ell}(z^{t-1}) \right) \end{cases} .$$

2.2 Analysis

Assumption 4 (Smoothness). Every $\nabla \ell_k$ is L-Lipschitz,

$$\forall z, z' \,,\, \|\nabla \ell_k(z) - \nabla \ell_k(z')\| \leqslant L\|z - z'\|.$$

Lemma 5. *If Assumption 4 is satisfied, then for all z, ξ, z', ξ',*

$$F(z', \xi') - F(z, \xi) - \left\langle \eth_{z'} F(z, \xi) \,,\, \begin{pmatrix} z' - z \\ \widetilde{\sigma}(\xi') - \widetilde{\sigma}(\xi) \end{pmatrix} \right\rangle \leqslant \frac{L}{2} \left\| z' - z \right\|^2 .$$

Proof. Since

$$\begin{cases} \sigma(\xi)^\intercal \ell(z') - \sigma(\xi)^\intercal \ell(z) - \sigma(\xi)^\intercal J_\ell(z)(z' - z) \leqslant \frac{L}{2}\|z' - z\|^2 \\ \sigma(\xi')^\intercal \ell(z) - \sigma(\xi')^\intercal \ell(z') - \sigma(\xi')^\intercal J_\ell(z')(z - z') \leqslant \frac{L}{2}\|z' - z\|^2 \end{cases} ,$$

we have by summing

$$\left\langle \nabla^G F(z, \xi) - \nabla^G F(z', \xi') \,,\, \nabla\Psi(z, \xi) - \nabla\Psi(z', \xi') \right\rangle \leqslant L\|z' - z\|^2.$$

Then, as in the proof of Lemma 3,

$$\begin{aligned} & F(z', \xi') - F(z, \xi) - \left\langle \eth_{z'} F(z, \xi) \,,\, \nabla\Psi(z', \xi') - \nabla\Psi(z, \xi) \right\rangle \\ &= \int_{\tau=0}^1 \frac{1}{\tau} \left\langle \widetilde{\nabla} F\left(z + \tau(z' - z), \widetilde{\sigma}^{-1}(\varphi(\tau)) \right) - \eth_{z'} F(z, \xi) \,,\, \tau(\nabla\Psi(z', \xi') - \nabla\Psi(z, \xi)) \right\rangle d\tau \\ &= \int_{\tau=0}^1 \frac{1}{\tau} \left\langle \nabla^G F\left(z + \tau(z' - z), \widetilde{\sigma}^{-1}(\varphi(\tau)) \right) - \nabla^G F(z, \xi) \,,\, \tau(\nabla\Psi(z', \xi') - \nabla\Psi(z, \xi)) \right\rangle d\tau \\ &\qquad\qquad \leqslant \frac{L}{2}\|z' - z\|^2. \end{aligned}$$

□

Theorem 6 (Regrets)**.** *Under Assumption 4 and if $\|\nabla \ell_k(z)\| \leqslant B$ $(\forall k, z)$, then HD with constant learning rates $\lambda_t = \lambda, \beta_t = \beta > 0$ satisfies*

(I) for all z^* such that $\|z^1 - z^*\| \leqslant D$,

$$\sum_{t=1}^T \left\langle \sigma(\xi^t) \,,\, \ell(z^t) - \ell(z^*) \right\rangle \leqslant \frac{D^2}{2\lambda} + \frac{B^2 T}{2}\lambda \,;$$

(II) for all ξ^* such that $D_{KL}(\sigma(\xi^*)\|\sigma(\xi^0)) \leqslant D'$,

$$\sum_{t=1}^{T} \left\langle \sigma(\xi^t) - \sigma(\xi^*)\,,\, \ell(z^t) \right\rangle \leqslant \frac{D'}{\beta} + \frac{LB^2T}{2}\lambda^2 - \frac{1}{\beta}\sum_{t=1}^{T} D_{KL}(\sigma(\xi^t)\|\sigma(\xi^{t-1})).$$

Proof. (I). We follow the proof of Theorem 2.1 in Bansal and Gupta [2019] with the potential function

$$\Phi_t = \frac{1}{2\lambda}\|z^t - z^*\|^2.$$

By Lemma 3,

$$F(z^t,\xi^t) - F(z^*,\xi^t) \leqslant \left\langle J_\ell(z^t)^\intercal \sigma(\xi^t)\,,\, z^t - z^* \right\rangle$$
$$\Rightarrow F(z^t,\xi^t) - F(z^*,\xi^t) + \Phi_{t+1} - \Phi_t \leqslant \frac{\lambda}{2}\|J_\ell(z^t)^\intercal\sigma(\xi^t)\|^2 \leqslant \frac{\lambda B^2}{2}.$$

Finally, by summing inequalities,

$$\sum_{t=1}^{T} F(z^t,\xi^t) - F(z^*,\xi^t) \leqslant \frac{D^2}{2\lambda} + \frac{B^2T}{2}\lambda.$$

(II). Let the potential function

$$\Phi_t = \frac{1}{\beta} D_{\mathrm{KL}}(\sigma(\xi^*)\|\sigma(\xi^t)).$$

By Lemma 5 applied to (z^{t-1},ξ^{t-1}) and (z^t,ξ^t), we obtain the descent property

$$F(z^t,\xi^t) - F(z^{t-1},\xi^{t-1}) \leqslant \left(-\lambda + \frac{L}{2}\lambda^2\right)\|J_\ell(z^{t-1})^\intercal\sigma(\xi^{t-1})\|^2 - \frac{1}{\beta}\left(D_{\mathrm{KL}}(\sigma(\xi^t)\|\sigma(\xi^{t-1})) + D_{\mathrm{KL}}(\sigma(\xi^{t-1})\|\sigma(\xi^t))\right), \quad (1)$$

and, by Lemma 3,

$$F(z^{t-1},\xi^{t-1}) - F(z^t,\xi^*) \leqslant \left\langle \eth_{z^t} F(z^{t-1},\xi^{t-1})\,,\, \begin{pmatrix} z^{t-1} - z^t \\ \widetilde{\sigma}(\xi^{t-1}) - \widetilde{\sigma}(\xi^*) \end{pmatrix} \right\rangle.$$

Adding these two inequalities provides

$$F(z^t,\xi^t) - F(z^t,\xi^*) + \Phi_t - \Phi_{t-1} \leqslant \frac{L}{2}\lambda^2B^2 - \frac{1}{\beta}D_{\mathrm{KL}}(\sigma(\xi^t)\|\sigma(\xi^{t-1}))$$

and we conclude by summing. □

3 Boolean satisfiability

3.1 Probabilistic 3-SAT

We consider a probabilistic 3-SAT problem over n independent Bernoulli random variables $X_1, \dots, X_n$ valued in $\{0,1\}$ and m clauses $C_1, \dots, C_m$. We assume that every clause C_i is a disjunction of three literals:

$$C_i = \bigcup_{j=1}^{3} \left(X_{\kappa_{i,j}} = b_{i,j}\right) \quad \text{with} \quad \kappa_{i,j} \in [n] := \{1, \dots, n\} \quad \text{and} \quad b_{i,j} \in \{0,1\}.$$

If all X_k's are deterministic, then maximizing $\prod_{i=1}^{m} \mathbb{P}(C_i)$ over the 2^n possible assignments $(X_1, \dots, X_n) \in \{0,1\}^n$, namely

$$\mathrm{SAT}(\kappa, b) := \max_{\text{deterministic } (X_k)_k \in \{0,1\}^n} \prod_{i=1}^{m} \mathbb{P}(C_i) \in \{0,1\}\,,$$

is exactly a standard 3-SAT problem in Conjunctive Normal Form (CNF). Alternatively, we can consider the random setting with probability of success parametrized as a sigmoid,

$$\mathbb{P}(X_k = 1) = \varsigma(z_k) := \frac{e^{z_k}}{1 + e^{z_k}}\,.$$

Clause probability. Let us assume that $j \neq j' \Rightarrow \kappa_{i,j} \neq \kappa_{i,j'}$. For $i \in [m]$, we have:

$$\begin{aligned}\mathbb{P}(C_i) = \mathbb{P}\left(\bigcup_{j=1}^{3} \left(X_{\kappa_{i,j}} = b_{i,j}\right)\right) &= 1 - \prod_{j=1}^{3} \varsigma((1-2b_{i,j}) z_{\kappa_{i,j}}) \\ &= \sum_{y \in \mathcal{Y}(b_i)} \frac{e^{\sum_{j=1}^{3} y_j \cdot z_{\kappa_{i,j}}}}{\prod_{j=1}^{3} (1 + e^{z_{\kappa_{i,j}}})} = \sum_{y \in \mathcal{Y}(b_i)} \underbrace{\prod_{j=1}^{3} \varsigma((2y_j - 1) z_{\kappa_{i,j}})}_{\text{log-concave}}\,, \quad (2)\end{aligned}$$

where $\mathcal{Y}(b_i) = \{0,1\}^3 \backslash \left\{ \begin{pmatrix} 1-b_{i,1} \\ 1-b_{i,2} \\ 1-b_{i,3} \end{pmatrix} \right\}$. Hence, the probability of each event C_i is a sum of 7 log-concave functions of $z = (z_1, \dots, z_n)$.

Regularized clause objective. From now on, we denote by σ the softargmax for $K = 6$. The negative log-likelihood of the i-th clause writes as

$$-\log(\mathbb{P}(C_i)) = -\log\left(\sum_{k=0}^{6} P_{i,k}(z)\right) \quad \text{with } P_{i,k} \text{ log-concave.}$$

After regularization we obtain the objective function

$$F_i(z, \Xi_i) = \sigma(\Xi_i)^\intercal \mathscr{L}_i(z) \quad \text{with} \quad \Xi_i \in \mathbb{R}^6 \quad \text{and} \quad \mathscr{L}_{i,k} = -\log(P_{i,k}).$$

3.2 Loss function

We define the average loss function across all clauses as

$$\boxed{F_{\mathrm{SAT}}(z,\Xi) = \frac{1}{m}\sum_{i=1}^{m} F_i(z,\Xi_i)}$$

where (z,Ξ) lives in the pseudo-Riemannian space characterized by the metric tensor[4]

$$G(\Xi) = \begin{pmatrix} I_n & 0 & \cdots & 0 \\ 0 & -\frac{1}{m}\mathbf{I}_\sigma(\Xi_1) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & -\frac{1}{m}\mathbf{I}_\sigma(\Xi_m) \end{pmatrix}.$$

The hourglass gradient vector admits the following expression:

$$\eth_{z'} F_{\mathrm{SAT}}(z,\Xi) = \begin{pmatrix} \frac{1}{m}\sum_{i=1}^{m}\sum_{k=0}^{6}\sigma_k(\Xi_i)\nabla\mathscr{L}_{i,k}(z) \\ 2\int_{\tau=0}^{1}\widetilde{\mathscr{L}_1}(z+\tau(z'-z))d\tau - \widetilde{\mathscr{L}_1}(z) \\ \vdots \\ 2\int_{\tau=0}^{1}\widetilde{\mathscr{L}_m}(z+\tau(z'-z))d\tau - \widetilde{\mathscr{L}_m}(z) \end{pmatrix}$$

where, for $\mathscr{L}_{i,k}$ associated to $y \in \mathcal{Y}(b_i)$,

$$\nabla\mathscr{L}_{i,k}(z) = \sum_{j=1}^{3}(1-2y_j)\varsigma((1-2y_j)z_{\kappa_{i,j}})\mathbf{e}_{\kappa_{i,j}}$$

with $\mathbf{e}_l \in \mathbb{R}^n$ the l-th "one-hot" basis vector, for any $l \in [n]$. Furthermore, the dilogarithm function $\mathrm{Li}_2(v) = -\int_0^v \frac{\log(1-u)}{u}\,du$ allows to compute the integrated losses:

$$\int_{\tau=0}^{1} -\log\varsigma(x+\tau(x'-x))\,d\tau = \frac{-\mathrm{Li}_2(-e^{-x'}) + \mathrm{Li}_2(-e^{-x})}{x'-x}\ . \tag{3}$$

3.3 Infimum

The infimum value of F_{SAT} is given by

$$\inf_{z,\Xi} F_{\mathrm{SAT}}(z,\Xi) = \min_{k_1,\ldots,k_m\in\{0,\ldots,6\}} \inf_z \frac{1}{m}\sum_{i=1}^{m}\mathscr{L}_{i,k_i}(z). \tag{4}$$

If the ground-truth is SAT, then $\inf F_{\mathrm{SAT}} = 0$. Otherwise, we have the following lower bound.

[4]justifies by showing that the pseudo-Riemannian gradient of $\frac{1}{m}\sum_{i=1}^{m} F_i(z,\Xi_i)$ remains Bregman monotone

Lemma 7. *If the instance is inherently UNSAT (i.e. SAT$(\kappa, b) = 0$), then*

$$\inf F_{SAT} \geqslant \frac{\log 4}{m}.$$

Proof. If $\mathrm{SAT}(\kappa, b) = 0$, then for any choice of $k_1, \ldots, k_m \in \{0, \ldots, 6\}$, there is at least an index $k \in [n]$ such that

$$\sum_{i=1}^{m} \mathscr{L}_{i,k_i}(z) \geqslant -\log(\varsigma(z_k)) - \log(1 - \varsigma(z_k)) \geqslant -\log \frac{1}{4},$$

which allows to conclude from Eq. (4). □

3.4 SubZero algorithm

1. Initialization: Choose initial point (z, Ξ)
2. Hourglass Descent: Repeat T times

$$\begin{pmatrix} z \\ \Xi \end{pmatrix} \leftarrow \begin{pmatrix} z \\ \Xi \end{pmatrix} - \begin{pmatrix} \lambda I_n & 0 \\ 0 & \beta I_{6m} \end{pmatrix} \eth_{z'} F_{\mathrm{SAT}}(z, \Xi)$$

$$\text{with} \quad z' = z - \frac{\lambda}{m} \sum_{i=1}^{m} \sum_{k=0}^{6} \sigma_k(\Xi_i) \nabla \mathscr{L}_{i,k}(z)$$

3. (i) If $F_{\mathrm{SAT}}(z, \Xi) < \frac{\log 4}{m}$ return 'SAT' and freeze assignment

$$\left(\frac{1 + \mathrm{sgn}(z_1)}{2}, \ldots, \frac{1 + \mathrm{sgn}(z_n)}{2} \right)$$

   (ii) Else return 'UNSAT'

References

Mastane Achab. Beyond log-concavity: Theory and algorithm for sum-log-concave optimization. *arXiv preprint arXiv:2309.15298*, 2023.

Mastane Achab. A bregman firmly nonexpansive proximal operator for baryconvex optimization. *arXiv preprint arXiv:2411.00928*, 2024.

Shun-ichi Amari. *Information geometry and its applications*. Springer, 2016.

Francis Bach. *Learning theory from first principles*. MIT press, 2024.

Nikhil Bansal and Anupam Gupta. Potential-function proofs for gradient methods. *Theory of Computing*, 15(1):1–32, 2019.

Armin Biere, Hans van Maaren, and Toby Walsh. *Handbook of satisfiability*. SAGE Publications Limited, 2009.

Sébastien Bubeck. Convex optimization: Algorithms and complexity. *Foundations and trends in Machine Learning*, 8(3-4):231–357, 2015.

Augustin Louis Cauchy. Méthode générale pour la résolution des systemes d'équations simultanées. *Comp. Rend. Sci. Paris*, 25(1847):536–538, 1847.

Arkadij Semenovič Nemirovskij and David Borisovich Yudin. Problem complexity and method efficiency in optimization. 1983.