Global Solutions to Non-Convex Functional Constrained Problems with Hidden Convexity

Ilyas Fatkhullin[1,2,3]
ilyas.fatkhullin@ai.ethz.ch

Niao He[1]
niao.he@inf.ethz.ch

Guanghui Lan[3]
george.lan@isye.gatech.edu

Florian Wolf[4*]
fwolf@caltech.edu

[1]Department of Computer Science, ETH Zürich, Switzerland
[2]ETH AI Center, Zürich, Switzerland
[3]H. Milton Stewart School of Industrial and Systems Engineering,
Georgia Institute of Technology, Atlanta, USA
[4]Computing & Mathematical Sciences Department,
California Institute of Technology, Pasadena, USA

Abstract

Constrained non-convex optimization is fundamentally challenging, as global solutions are generally intractable and constraint qualifications may not hold. However, in many applications, including safe policy optimization in control and reinforcement learning, such problems possess hidden convexity, meaning they can be reformulated as convex programs via a nonlinear invertible transformation. Typically such transformations are implicit or unknown, making the direct link with the convex program impossible. On the other hand, (sub)-gradients with respect to the original variables are often accessible or can be easily estimated, which motivates algorithms that operate directly in the original (non-convex) problem space using a standard (sub)-gradient oracle. In this work, we develop the first algorithms that provably solve such non-convex problems to global minima. Surprisingly, despite non-convexity, our methodology does not require constraint qualifications and achieves complexities matching those for unconstrained hidden convex optimization.

1 Introduction

Non-convex constrained optimization problems (with possibly non-smooth objectives and constraints) arise frequently in many modern applications. In this work, we study a sub-class of non-convex problems of the form

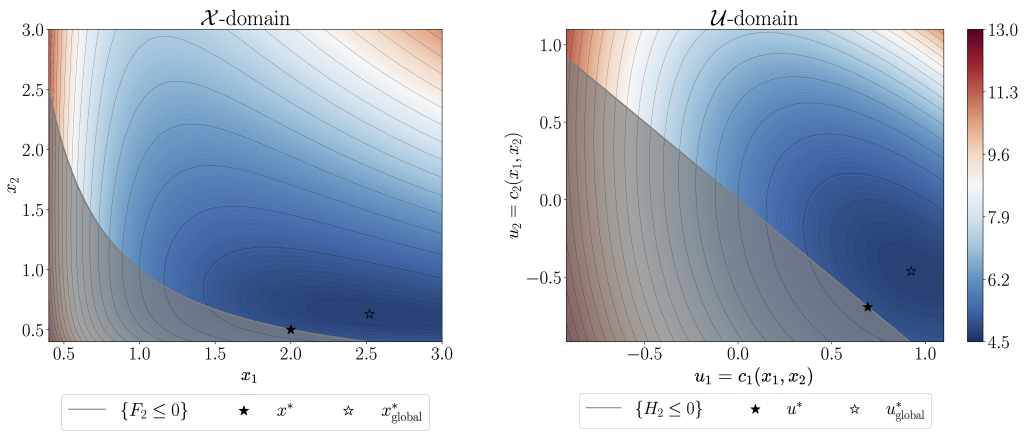
$$\min_{x\in\mathcal{X}} F_1(x), \qquad \text{s.t. } F_2(x)\le 0, \tag{1}$$

where $\mathcal{X}\subset\mathbb{R}^d$ is a closed convex set, and F_1, F_2 are possibly *non-convex* with respect to variable x. A central running assumption in this work is that problem (1) admits a *convex reformulation*

$$\min_{u\in\mathcal{U}} H_1(u), \qquad \text{s.t. } H_2(u)\le 0 \tag{2}$$
$$\textit{via variable change:} \quad u=c(x),$$

*Work carried out while at ETH Zürich.

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: Constrained Optimization for Machine Learning (COML).



(a) **Non-Convex Formulation.** Level sets of F_1 and the feasible set $\{F_2 \leq 0\}$ in the $\mathcal{X}$-domain.

(b) **Convex Reformulation.** Level sets of H_1 and the feasible set $\{H_2 \leq 0\}$ in the $\mathcal{U}$-domain.

Figure 1: **Constrained Geometric Programming** problem [Eck80; BKV+07; Xia20] with $F_1(x) \coloneqq x_1 \cdot x_2 + \frac{4}{x_1} + \frac{1}{x_2}$ constrained to the set $\{F_2 \leq 0\} \coloneqq \{x \in \mathcal{X} \mid F_2(x) \leq 0\}$ with $F_2(x) \coloneqq x_1 \cdot x_2 - 1$ is an example of a smooth hidden convex problem under the transformation $c(x) \coloneqq (\log x_1, \log x_2)^\top$. We use the notation x^*_{global} to denote the global optimum of $\min_x F_1(x)$ without constraints and x^* to denote the minimizer under constraints, with u^*_{global} and u^* respectively in $\mathcal{U}$. The gray region denotes the feasible set and the objective value of F_1 is illustrated in color.

where H_1, H_2 are convex functions defined over a closed convex set $\mathcal{U} \subset \mathbb{R}^d$, and $c : \mathcal{X} \to \mathcal{U}$ is an invertible map (with c^{-1} denoting its inverse). This property, often referred to as *hidden convexity*, appears in diverse applications, including policy optimization in optimal control [SF21; ADL+19; FP21; ZPT23; WZ25] and reinforcement learning [ZKB+20; BFH23; YGL+24], variational inference [WG24; SFH25], generative models [KPB21], supply chain and revenue management [FS18; CHH+25], geometric programming [Xia20], neural network training [Bac17; WLP22], and non-monotone games [VFP21; MSG+22; SVM+23; DVL+25], to name a few. A simple instance of a hidden convex constrained problem is illustrated in Figure 1, while Appendix A provides further details on a key motivating example in the context of constrained reinforcement learning.

In this work, we analyze the Proximal Point Method (PPM) for hidden convex problems under hidden convex constraints, establishing a last iterate convergence rate of $\widetilde{\mathcal{O}}(\varepsilon^{-1})$ proximal point evaluations to achieve $(\varepsilon, \varepsilon)$-approximate global minima, i.e., $F_1(x^{(N)}) - F_1^* \leq \varepsilon$, $F_2(x^{(N)}) \leq \varepsilon$. Next, we carefully approximate the PPM using existing algorithms for strongly convex constrained optimization. In particular, in non-smooth setting, we use switching sub-gradient method as inner solver to derive $\widetilde{\mathcal{O}}(\varepsilon^{-3})$ oracle complexity without any constraint qualification assumptions (CQs). In smooth setting, we use accelerated constrained gradient method to further improve the above result to $\widetilde{\mathcal{O}}(\varepsilon^{-2})$ (and $\widetilde{\mathcal{O}}(\theta^{-1}\varepsilon^{-1})$) oracle complexity without any CQs (and with θ-Slater condition). In addition, in the non-smooth setting our gradient complexity improves the previously known ones for simpler unconstrained hidden convex problems from $\widetilde{\mathcal{O}}(\varepsilon^{-6})$ to $\widetilde{\mathcal{O}}(\varepsilon^{-3})$ using a different algorithm. See Table 1 for a summary.

Contribution. Constrained proximal point method (PPM) analysis with last iterate convergence results without requiring any constraint qualification assumptions.

2 Notation, assumption and problem formulation

A map $c : \mathcal{X} \to \mathcal{U}$ is invertible if there exists $c^{-1} : \mathcal{U} \to \mathcal{X}$ such that $c^{-1}(c(x)) = x$ and $c(c^{-1}(u)) = u$ for all $x \in \mathcal{X}$, $u \in \mathcal{U}$. A function $F_i : \mathcal{X} \to \mathbb{R}$ is ρ-weakly convex (ρ-WC) if for all $y \in \mathcal{X}$, the function $F_{i,\rho}(x, y) \coloneqq F_i(x) + \frac{\rho}{2}\|x - y\|^2$ is convex in $x \in \mathcal{X}$. The (Fréchet) subdifferential at $x \in \mathcal{X}$ is $\partial F_i(x) \coloneqq \{g_i \in \mathbb{R}^d \mid F_i(y) \geq F_i(x) + \langle g_i, y - x\rangle + o(\|y - x\|), \forall y \in \mathbb{R}^d\}$,

Smoothness	Setting	Method	Complexity
Non-smooth	$F_2(\cdot) \equiv 0$	Sub-gradient Method (SM)	$\widetilde{\mathcal{O}}\left(\varepsilon^{-6}\right)$ [FHH24]
	$F_2(\cdot) \equiv 0$	PP + SM	$\widetilde{\mathcal{O}}\left(\varepsilon^{-3}\right)$
	$F_2(\cdot) \not\equiv 0$	PP + Switching Sub-gradient	$\widetilde{\mathcal{O}}\left(\varepsilon^{-3}\right)$
L-smooth	$F_2(\cdot) \equiv 0$	Gradient Descent	$\widetilde{\mathcal{O}}\left(\varepsilon^{-1}\right)$ [ZKB+20; FHH24]
	$F_2(\cdot) \not\equiv 0$	PP + ACGD	$\widetilde{\mathcal{O}}\left(\varepsilon^{-2}\right)$ or $\widetilde{\mathcal{O}}\left(\theta^{-1}\varepsilon^{-1}\right)$ (θ is Slater's gap)

Table 1: Summary of **total** *(sub-)gradient* and *function evaluation* **complexities** under *hidden convexity*. The "Setting" column distinguishes between unconstrained ($F_2(\cdot) \equiv 0$) and constrained ($F_2(\cdot) \not\equiv 0$) problems. The third column reports the number of (sub-)gradient (and function, when $F_2(\cdot) \not\equiv 0$) evaluations to find a point $x^{(N)}$ such that $F_1(x^{(N)}) - F_1^* \leq \varepsilon$, $F_2(x^{(N)}) \leq \varepsilon$. The complexity of SM in [FHH24] is stated in terms of the Moreau envelope and suffers a loss in complexity when translated to the original objective in the non-smooth setting, see the discussion after Corollary 1 therein. "PP" stands for Proximal Point method, Algorithm 1, "ACGD" refers to Accelerated Constrained Gradient Descent.

and its elements $g_i \in \partial F_i(x)$ are called subgradients. A differentiable function F_i is L-smooth on $\mathcal{X} \subset \mathbb{R}^d$ if its gradient is L-Lipschitz, i.e., $\|\nabla F_i(x) - \nabla F_i(y)\| \leq L\|x - y\|$, for all $x, y \in \mathcal{X}$.

Assumption 1. *We first make the following (standard) assumptions:*

1. *The functions F_1, F_2 are ρ-weakly convex.*
2. *The functions F_1, F_2 satisfy for all $x \in \mathcal{X}$ that $\partial F_1(x) \neq \emptyset$, $\partial F_2(x) \neq \emptyset$ on $\mathcal{X}$, and the norms of the subgradients are uniformly bounded by $\|g_1\| \leq G_{F_1}$, $\|g_2\| \leq G_{F_2}$ for all $x \in \mathcal{X}$, $g_1 \in \partial F_1(x)$ and $g_2 \in \partial F_2(x)$ respectively. We define $G \coloneqq \max\{G_{F_1}, G_{F_2}\}$.*
3. *The domain $\mathcal{U}$ has bounded diameter $\mathcal{D}_{\mathcal{U}} > 0$.*

Definition 1 (Hidden Convexity)**.** *The problem (1) is called* hidden convex *with modulus $\mu_c > 0$, if its components satisfy the following underlying conditions.*

1. *The domain $\mathcal{U} = c(\mathcal{X})$ is convex, the functions $H_1, H_2 : \mathcal{U} \to \mathbb{R}$ are convex, i.e. satisfy for $i = 1, 2$ and for all $u, v \in \mathcal{U}$ and $\lambda \in [0, 1]$*

$$H_i((1-\lambda)u + \lambda v) \leq (1-\lambda)H_i(u) + \lambda H_i(v). \tag{HC-1}$$

   Additionally, we assume (1) admits a solution $u^* \in \mathcal{U}$ with its corresponding objective function value $F_1^* \coloneqq H_1(u^*) = F_1(c^{-1}(u^*))$.

2. *The map $c : \mathcal{X} \to \mathcal{U}$ is invertible and there exists a $\mu_c > 0$ such that for all $x, y \in \mathcal{X}$ it holds*

$$\|c(x) - c(y)\| \geq \mu_c\|x - y\|. \tag{HC-2}$$

Note that the condition (HC-2) along with Assumption 1 (Item 3) imply that the domain $\mathcal{X}$ has bounded diameter $\mathcal{D}_{\mathcal{X}} \leq \frac{1}{\mu_c}\mathcal{D}_{\mathcal{U}}$.

3 Proximal Point Method (PPM) under Hidden Convexity

We want to analyze the inexact Proximal Point Method (IPPM), which solves in each iteration $k \in [N]$ the following (strongly convex) problem

$$\begin{aligned} x^{(k+1)} \approx \hat{x}^{(k+1)} \coloneqq \arg\min_{x \in \mathcal{X}} \ & \varphi_1^{(k)}(x) \coloneqq F_1(x) + \frac{\hat{\rho}}{2}\|x - x^{(k)}\|^2, \\ \text{s.t. } & \varphi_2^{(k)}(x) \coloneqq F_2(x) + \frac{\hat{\rho}}{2}\|x - x^{(k)}\|^2 \leq \tau. \end{aligned} \tag{HC-IPPM}$$

Here, $\hat{x}^{(k+1)}$ denotes the exact solution to the subproblem, and $x^{(k+1)}$ is an approximate solution.

Algorithm 1 IPPM$(F_1, F_2, x^{(0)}, \varepsilon, \tau, N, \hat{\rho})$
Inexact Proximal Point Method for (1)

1: **Input:** Objective F_1, constraint F_2, initial point $x^{(0)} \in \mathcal{X} \cap \{F_2(\,\cdot\,) \le \tau\}$, accuracy ε, constraint violation budget τ, outer loops N, inner (feasible) algorithm $\mathcal{A}$, regularization parameter $\hat{\rho} > \rho$
2: **for** iteration $k = 0, 1, \ldots, N-1$ **do**
3: Define

$$\varphi_i^{(k)}(x) := F_i(x) + \frac{\hat{\rho}}{2}\|x - x^{(k)}\|^2,\ i = 1, 2,\ x \in \mathcal{X}$$

4: Find $x^{(k+1)} \leftarrow \mathcal{A}_{\varepsilon_{\text{in}}}(\varphi_1^{(k)}, \varphi_2^{(k)}, x^{(k)}, T_{\text{in}}, \tau)$ as an approximate solution to (HC-IPPM)
5: **end for**
6: **Return:** $x^{(N)}$

Suppose the algorithm $\mathcal{A}_{\varepsilon_{\text{in}}}$ solves (HC-IPPM) $(\varepsilon_{\text{in}}, 0)$-optimally for any target precision $\varepsilon_{\text{in}} > 0$, i.e.

$$\begin{aligned} \varphi_1^{(k)}(x^{(k+1)}) - \varphi_1^{(k)}(\hat{x}^{(k+1)}) &\le \varepsilon_{\text{in}}, \\ \varphi_2^{(k)}(x^{(k+1)}) - \tau &\le 0. \end{aligned} \tag{IPPM-Feas}$$

Theorem 1 (Inexact PPM). *Assume that (1) is hidden convex and Assumption 1 holds. Let $x^{(0)}$ be τ–feasible for (1) and algorithm $\mathcal{A}_{\varepsilon_{\text{in}}}$ initialized with a feasible point $x^{(k)}$ outputs a point $x^{(k+1)}$ satisfying (IPPM-Feas) after $T_{\text{in}} = T_{\text{in}}(\varepsilon_{\text{in}})$ (sub)-gradient and function evaluations. Given a lifting parameter $\hat{\rho} > \rho$ and a desired tolerance $\varepsilon > 0$ for the optimality gap, assume that $\varepsilon \le \frac{3\hat{\rho}\mathcal{D}_{\mathcal{U}}^2}{2\mu_c^2}$ and $\tau \le \frac{\hat{\rho}\mathcal{D}_{\mathcal{U}}^2}{2\mu_c^2}$ hold. Then setting $\hat{\rho} := 2\rho$ and $\varepsilon_{\text{in}} \le {}^{(\alpha\,\varepsilon)}/_{3}$, the last iterate of Algorithm 1 satisfies*

$$F_1(x^{(N)}) - F_1^* \le \varepsilon, \quad F_2(x^{(N)}) \le \tau \tag{IPPM-Opt}$$

after $N \ge \max\left\{\frac{3\rho\mathcal{D}_{\mathcal{U}}^2}{\mu_c^2\varepsilon}, \frac{2\rho\mathcal{D}_{\mathcal{U}}^2}{\mu_c^2\tau}\right\} \cdot \log\left(\frac{3\Delta_0}{\varepsilon}\right)$ iterations, where $\Delta_0 := F_1(x^{(0)}) - F_1^*$. The total oracle complexity is given by

$$T_{\text{tot}} \ge N \cdot T_{\text{in}}(\varepsilon_{\text{in}}) = \max\left\{\frac{3\rho\mathcal{D}_{\mathcal{U}}^2}{\mu_c^2\varepsilon}, \frac{2\rho\mathcal{D}_{\mathcal{U}}^2}{\mu_c^2\tau}\right\} \cdot \log\left(\frac{3\Delta_0}{\varepsilon}\right) \cdot T_{\text{in}}(\varepsilon_{\text{in}}).$$

We compute the total oracle complexity bounds, including the complexity $T_{\text{in}}(\varepsilon_{\text{in}})$ to solve the inner IPPM subproblem, in Table 1. We use Switching Sub-gradient [Pol67; LZ20; JG25] and Accelerated Constrained Gradient Descent [ZL22] methods for inner solvers.

Proof sketch:

1. *Initialization.* Assume $x^{(0)}$ is τ-feasible for (1), i.e., $F_2(x^{(0)}) \le \tau$. If not, a simple gradient method on F_2 finds such a point in $\widetilde{\mathcal{O}}(1/\tau)$ gradient evaluations in the smooth case or $\widetilde{\mathcal{O}}\big(1/\tau^3\big)$ in the non-smooth case; this does not change the overall complexity.
2. *Feasibility preservation.* If $x^{(k)}$ is τ-feasible for (1), it is (trivially) feasible for (HC-IPPM). Running a *feasible* inner method on (HC-IPPM) from $x^{(k)}$ yields $\varphi_2^{(k)}(x^{(k)}) \le \tau$ and hence $F_2(x^{(k)}) \le \tau$. Thus all outer iterates remain τ-feasible for (1).
3. *Slater points for subproblems.* Define $x_\alpha^{(k)} := c^{-1}\big((1-\alpha)c(x^{(k)}) + \alpha c(x^\star)\big)$. For sufficiently small α (relative to τ), the point $x_\alpha^{(k)}$ is *strictly feasible* for (HC-IPPM). This proves the Slater condition for (HC-IPPM) at each iteration $k \ge 0$ and allows us to apply a feasible method to solve this subproblem.
4. *Optimality Improvement.* It remains to build a PPM recursion similar to analysis in [FHH24] to guarantee the improvement in F_1, up to errors from inexact inner solvers.

Constructing the Slater points in step 3. of the above analysis is the key ingredient allowing convergence of IPPM in Theorem 1. Now we formalize our "HC–Slater lemma", which verifies the PPM subproblems (HC-IPPM) satisfy the Slater condition under a suitable reference point $x^{(k)}$.

Lemma 1 (HC–Slater lemma)**.** *Assume that* (1) *is hidden convex, Assumption 1.3 holds and* $x^{(k)}$ *is* τ*–feasible for* (1)*. Then*

1. (HC-IPPM) *satisfies* $\frac{\alpha\tau}{2}$*–Slater condition with* $\alpha \leq \min\left\{1, \frac{\mu_c^2\tau}{\hat{\rho}\mathcal{D}_{\mathcal{U}}^2}\right\}$.
2. *If, additionally,* (1) *satisfies* θ*–Slater condition, then* (HC-IPPM) *satisfies* $\frac{\beta\theta}{2}$*–Slater condition with* $\beta \leq \min\left\{1, \frac{\mu_c^2\theta}{\hat{\rho}\mathcal{D}_{\mathcal{U}}^2}\right\}$.

Lemma 1 says that regardless whether (1) satisfies Slater condition, the Slater point for PPM subproblem (HC-IPPM) always exists. The proof of this result is deferred to Appendix B.2.

References

[ADL+19] J. Anderson et al. *System Level Synthesis*. Apr. 2019. arXiv: 1904.01634 (cit. on p. 2).

[Bac17] F. Bach. "Breaking the Curse of Dimensionality with Convex Neural Networks". In: *Journal of Machine Learning Research* 18.19 (2017), pp. 1–53 (cit. on p. 2).

[BFH23] A. Barakat, I. Fatkhullin, and N. He. "Reinforcement Learning with General Utilities: Simpler Variance Reduction and Large State-Action Space". In: *Proceedings of the 40th International Conference on Machine Learning*. PMLR, July 2023, pp. 1753–1800 (cit. on p. 2).

[BKV+07] S. Boyd et al. "A Tutorial on Geometric Programming". In: *Optimization and Engineering* 8.1 (Mar. 2007), pp. 67–127 (cit. on p. 2).

[CHH+25] X. Chen et al. "Efficient Algorithms for a Class of Stochastic Hidden Convex Optimization and Its Applications in Network Revenue Management". In: *Operations Research* 73.2 (2025), pp. 704–719. arXiv: 2205.01774 (cit. on p. 2).

[DVL+25] R. D'Orazio et al. "Solving Hidden Monotone Variational Inequalities with Surrogate Losses". In: *The Thirteenth International Conference on Learning Representations*. 2025 (cit. on p. 2).

[Eck80] J. G. Ecker. "Geometric Programming: Methods, Computations and Applications". In: *SIAM Review* 22.3 (July 1980), pp. 338–362 (cit. on p. 2).

[FHH24] I. Fatkhullin, N. He, and Y. Hu. *Stochastic Optimization under Hidden Convexity*. Nov. 2024. arXiv: 2401.00108 (cit. on pp. 3, 4, 8).

[FP21] I. Fatkhullin and B. T. Polyak. "Optimizing Static Linear Feedback: Gradient Method". In: *SIAM J. Control. Optim.* 59.5 (2021), pp. 3887–3911. arXiv: 2004.09875 (cit. on p. 2).

[FS18] Q. Feng and J. G. Shanthikumar. "Supply and Demand Functions in Inventory Models". In: *Operations Research* 66.1 (2018), pp. 77–91 (cit. on p. 2).

[GPL+21] M. Geist et al. "Concave utility reinforcement learning: The mean-field game viewpoint". In: *arXiv preprint arXiv:2106.03787* (2021) (cit. on p. 8).

[JG25] Z. Jia and B. Grimmer. "First-Order Methods for Nonsmooth Nonconvex Functional Constrained Optimization with or without Slater Points". In: *SIAM J. Optim.* 35.2 (2025), pp. 1300–1329 (cit. on p. 4).

[KPB21] I. Kobyzev, S. J. Prince, and M. A. Brubaker. "Normalizing Flows: An Introduction and Review of Current Methods". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 43.11 (Nov. 2021), pp. 3964–3979 (cit. on p. 2).

[LZ20] G. Lan and Z. Zhou. "Algorithms for Stochastic Optimization with Function or Expectation Constraints". In: *Computational Optimization and Applications* 76.2 (June 2020), pp. 461–498 (cit. on p. 4).

[MDD+22] M. Mutti et al. "Challenging common assumptions in convex reinforcement learning". In: *Advances in Neural Information Processing Systems* 35 (2022), pp. 4489–4502 (cit. on p. 8).

[MSG+22] A. Mladenovic et al. "Generalized Natural Gradient Flows in Hidden Convex-Concave Games and GANs". In: *International Conference on Learning Representations*. 2022 (cit. on p. 2).

[Pol67] B. Polyak. "A General Method for Solving Extremum Problems". In: *Soviet Mathematics. Doklady* 8 (Jan. 1967) (cit. on p. 4).

[SB18] R. S. Sutton and A. G. Barto. *Reinforcement Learning: An Introduction*. Second edition. Adaptive Computation and Machine Learning Series. Cambridge, Massachusetts: The MIT Press, 2018 (cit. on p. 8).

[SF21] Y. Sun and M. Fazel. "Learning Optimal Controllers by Policy Gradient: Global Optimality via Convex Parameterization". In: *2021 60th IEEE Conference on Decision and Control (CDC)*. Austin, TX, USA: IEEE Press, Dec. 2021, pp. 4576–4581 (cit. on p. 2).

[SFH25] F. Sun, I. Fatkhullin, and N. He. "Natural Gradient VI: Guarantees for Non-Conjugate Models". In: *arXiv preprint arXiv:2510.19163* (2025) (cit. on p. 2).

[SVM+23] I. Sakos et al. "Exploiting Hidden Structures in Non-Convex Games for Convergence to Nash Equilibrium". In: *Advances in Neural Information Processing Systems*. Vol. 36. 2023, pp. 66979–67006 (cit. on p. 2).

[VFP21] E.-V. Vlatakis-Gkaragkounis, L. Flokas, and G. Piliouras. “Solving Min-Max Optimization with Hidden Structure via Gradient Descent Ascent”. In: *Advances in Neural Information Processing Systems*. Vol. 34. Curran Associates, Inc., 2021, pp. 2373–2386 (cit. on p. 2).

[WG24] K. Wu and J. R. Gardner. “Understanding Stochastic Natural Gradient Variational Inference”. In: *International Conference on Machine Learning*. PMLR. 2024, pp. 53398–53421 (cit. on p. 2).

[WLP22] Y. Wang, J. Lacotte, and M. Pilanci. “The Hidden Convex Optimization Landscape of Regularized Two-Layer Relu Networks: An Exact Characterization of Optimal Solutions”. In: *International Conference on Learning Representations*. 2022. arXiv: 2006.05900 (cit. on p. 2).

[WZ25] Y. Watanabe and Y. Zheng. *Revisiting Strong Duality, Hidden Convexity, and Gradient Dominance in the Linear Quadratic Regulator*. Mar. 2025. arXiv: 2503.10964 (cit. on p. 2).

[Xia20] Y. Xia. “A Survey of Hidden Convex Optimization”. In: *Journal of the Operations Research Society of China* 8.1 (Mar. 2020), pp. 1–28 (cit. on p. 2).

[YGL+24] D. Ying et al. *Policy-Based Primal-Dual Methods for Concave CMDP with Variance Reduction*. May 2024. arXiv: 2205.10715 (cit. on p. 2).

[ZKB+20] J. Zhang et al. “Variational Policy Gradient Method for Reinforcement Learning with General Utilities”. In: *Advances in Neural Information Processing Systems*. Vol. 33. 2020, pp. 4572–4583 (cit. on pp. 2, 3, 8).

[ZL22] Z. Zhang and G. Lan. *Solving Convex Smooth Function Constrained Optimization Is Almost As Easy As Unconstrained Optimization*. Nov. 2022. arXiv: 2210.05807 (cit. on p. 4).

[ZOD+21] T. Zahavy et al. “Reward is enough for convex mdps”. In: *Advances in Neural Information Processing Systems* 34 (2021), pp. 25746–25759 (cit. on p. 8).

[ZPT23] Y. Zheng, C.-f. Pai, and Y. Tang. *Benign Nonconvex Landscapes in Optimal and Robust Control, Part I: Global Optimality*. Dec. 2023. arXiv: 2312.15332 (cit. on p. 2).

A Detailed Motivating Example: Convex Constrained Markov Decision Process (CCMPD) [ZKB+20]

Convex reinforcement learning (RL) under convex constraints generalizes the classical (constrained) RL setting. Based on a discounted constrained Markov Decision Process (CMDP) of the form $\mathcal{M}(\mathcal{S}, \mathcal{A}, \mathbb{P}, \mu_0, \gamma, r, c)$, where $\mathcal{S}$ and $\mathcal{A}$ denote the (finite) state and action spaces respectively, $\mathbb{P} : \mathcal{S} \times \mathcal{A} \to \Delta(\mathcal{S})$ represents the state-action transition probability kernel, μ_0 is the initial state distribution and $\gamma \in (0, 1)$ is the discount factor. Based on the reward $r : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$ and the penalty cost $c : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$ classical RL is formulated in finding an optimal stationary policy $\pi : \Delta(\mathcal{A})^{|\mathcal{S}|} \to \Delta(\mathcal{A})$ by maximizing the reward function while satisfying the constraint on the cost function, i.e. formally

$$\begin{aligned} \min_{\pi \in \Pi} \; & F_1(\pi) \coloneqq -\mathbb{E}_{s_0 \sim \mu_0, \pi} \left[\sum_{h=0}^{\infty} \gamma^h r(s_h, a_h) \right], \\ \text{s.t. } & F_2(\pi) \coloneqq \mathbb{E}_{s_0 \sim \mu_0, \pi} \left[\sum_{h=0}^{\infty} \gamma^h c(s_h, a_h) \right] \leq 0, \end{aligned} \tag{CMDP}$$

where $\Pi \coloneqq \Delta(\mathcal{A})^{|\mathcal{S}|}$ is the set of all stationary policies. This set is the product of simplices, which admits an efficient projection.

Given a policy π, at each time $h \in \mathbb{N}$, the agent is in a state s_h and chooses an action $a_h \sim \pi(\cdot\,|s_h)$, resulting in a transition $s_{h+1} \sim \mathbb{P}(\cdot\,|s_h, a_h)$. With $\mathbb{P}_{\mu_0, \pi}$ we denote the induced probability distribution of the Markov chain $(s_h, a_h)_{h \in \mathbb{N}}$ with an initial state distribution μ_0. Under a transformation via the *state-action occupancy measure* [SB18], defined by

$$\lambda^\pi : \mathcal{S} \times \mathcal{A} \to (0, 1), \; \lambda^\pi(s, a) \coloneqq \sum_{h=0}^{\infty} \gamma^h \mathbb{P}_{\mu_0, \pi}(s_h = s, a_h = a), \tag{3}$$

the classical constrained RL problem becomes linear in the objective and the constraint, i.e. (CMDP) is equivalent to $\min_{\lambda \in \mathcal{U}} \langle r, \lambda^\pi \rangle$ s.t. $\langle c, \lambda^\pi \rangle \leq 0$, where $\mathcal{U} \coloneqq \{\lambda^\pi \mid \pi \in \Pi\}$. Convex constrained RL generalizes this optimization problem to

$$\begin{aligned} \min_{\pi \in \Pi} \; & F_1(\pi) \coloneqq H_1(\lambda^\pi), \\ \text{s.t. } & F_2(\pi) \coloneqq H_2(\lambda^\pi) \leq 0, \end{aligned} \tag{CCMDP}$$

where the utility functions $H_1, H_2 : \mathcal{U} \to \mathbb{R}$ are convex functions in λ^π, but the resulting optimization problem over the policy space Π is non-convex. Beyond the standard (CMDP), popular examples of (CCMDP) encompass safe exploration, i.e. H_1 is the negative entropy, under safety constraints, e.g. staying close to an experts trajectory $H_2(\lambda^\pi) \coloneqq \|\lambda^\pi - \lambda^{\pi_{\text{exp}}}\|$. Other instances include safe apprenticeship learning and safe learning, see e.g., [GPL+21; MDD+22; ZOD+21] for details.

The problem (CCMDP) is hidden convex by construction with $\mathcal{X} = \Pi$ and $c(x) \coloneqq \lambda^\pi$, (with $x = \pi$). As shown in [ZKB+20], the constant μ_c can be estimated under mild assumptions on the initial distribution μ_0. Note that in convex RL, we can control λ^π only implicitly by changing the policy π, i.e. via the policy gradient theorem, and thus, the exact computation of the transformation map and its inverse would require the knowledge of the state-action transition probability kernel and can be either computationally expensive or even intractable.

B Missing Proofs

B.1 Key inequalities for analysis under hidden convexity

Proposition 1 ([FHH24] Prop. 3)**.** *Let $F_i(\cdot)$, $i = 1, 2$, be hidden convex with $\mu_c > 0$. For any $\alpha \in [0, 1]$ and $x, y \in \mathcal{X}$, define $x_\alpha \coloneqq c^{-1}((1-\alpha)c(x) + \alpha c(y))$, then, for $i = 1, 2$, the following functional inequality*

$$F_i(x_\alpha) \leq (1 - \alpha) F_i(x) + \alpha F_i(y), \tag{HC-FI}$$

and norm inequality

$$\|x_\alpha - x\| \leq \frac{\alpha}{\mu_c} \|c(x) - c(y)\|. \tag{HC-NI}$$

hold.

B.2 Proof of HC–Slater Lemma

Proof. **Part 1.** Fix an iteration index $k \in [N]$, and for any $\alpha \in [0,1]$ define

$$x_\alpha^{(k)} := c^{-1}\left((1-\alpha)c(x^{(k)}) + \alpha c(x^*)\right).$$

We will show that $x_\alpha^{(k)}$ is a Slater point for subproblem (HC-IPPM). Indeed,

$$\begin{aligned}
F_2(x_\alpha^{(k)}) &+ \frac{\hat\rho}{2}\|x_\alpha^{(k)} - x^{(k)}\|^2 - \tau \\
&\overset{(i)}{\leq} (1-\alpha)F_2(x^{(k)}) + \alpha F_2(x^*) + \frac{\hat\rho}{2}\|x_\alpha^{(k)} - x^{(k)}\|^2 - \tau \\
&\overset{(ii)}{\leq} (1-\alpha)\tau + \alpha\cdot 0 + \frac{\hat\rho}{2}\|x_\alpha^{(k)} - x^{(k)}\|^2 - \tau \\
&= -\alpha\tau + \frac{\hat\rho}{2}\|c^{-1}\left((1-\alpha)c(x^{(k)}) + \alpha c(x^*)\right) - c^{-1}(c(x^{(k)}))\|^2 \\
&\overset{(iii)}{\leq} -\alpha\tau + \frac{\hat\rho}{2}\frac{\alpha^2}{\mu_c^2}\|c(x^{(k)}) - c(x^*)\|^2 \\
&\overset{(iv)}{\leq} -\alpha\tau + \frac{\hat\rho}{2}\mathcal{D}_{\mathcal{U}}^2\frac{\alpha^2}{\mu_c^2} \leq -\frac{\alpha\,\tau}{2} < 0,
\end{aligned} \tag{4}$$

where we used (HC-FI) for F_2 in (i), τ-feasibility of $x^{(k)}$ in (ii), (HC-NI) in (iii), as well as the boundedness of $\mathcal{U}$ domain in (iv). The last inequality follows by the choice of α.

Part 2. Let $\bar y \in \mathcal{X}$ be a θ–Slater point of (1), and for any $k \in [N]$, $\beta \in [0,1]$ define

$$\bar y_\beta^{(k)} := c^{-1}\left((1-\beta)c(x^{(k)}) + \beta c(\bar y)\right).$$

We will show that $\bar y_\beta^{(k)}$ is a Slater point for subproblem (HC-IPPM). Indeed,

$$\begin{aligned}
F_2(\bar y_\beta^{(k)}) &+ \frac{\hat\rho}{2}\|\bar y_\beta^{(k)} - x^{(k)}\|^2 - \tau \\
&\overset{(i)}{\leq} (1-\beta)F_2(x^{(k)}) + \beta F_2(\bar y) + \frac{\hat\rho}{2}\|\bar y_\beta^{(k)} - x^{(k)}\|^2 - \tau \\
&\overset{(ii)}{\leq} (1-\beta)\tau - \beta\cdot\theta + \frac{\hat\rho}{2}\|\bar y_\beta^{(k)} - x^{(k)}\|^2 - \tau \\
&\overset{(iii)}{\leq} -\beta\theta + \frac{\hat\rho}{2}\|c^{-1}\left((1-\beta)c(x^{(k)}) + \beta c(\bar y)\right) - c^{-1}(c(x^{(k)}))\|^2 \\
&\overset{(iv)}{\leq} -\beta\theta + \frac{\hat\rho}{2}\frac{\beta^2}{\mu_c^2}\|c(x^{(k)}) - c(\bar y)\|^2 \\
&\overset{(v)}{\leq} -\beta\theta + \frac{\hat\rho}{2}\mathcal{D}_{\mathcal{U}}^2\frac{\beta^2}{\mu_c^2} \leq -\frac{\beta\,\theta}{2} < 0,
\end{aligned} \tag{5}$$

where we used (HC-FI) for F_2 in (i), θ–Slater point in (ii), $(1-\beta)\tau \leq \tau$ since $\beta \leq 1$ in (iii), (HC-NI) in (iv), as well as the boundedness of $\mathcal{U}$ domain in (v). The last inequality follows by the choice of β. □