SURVEYS TEMPLATE

The Convexity Zoo: A Taxonomy of Function Classes in Optimization

Abbas Khademi[a]

[a]School of Mathematics, Institute for Research in Fundamental Sciences (IPM), P.O.Box 19395-5746, Tehran, Iran

ARTICLE HISTORY
Compiled January 2, 2026

ABSTRACT
The tractability of optimization problems depends critically on structural properties of the objective function. Convexity guarantees global optimality of local solutions and enables polynomial-time algorithms under mild assumptions, but many problems arising in modern applications—particularly in machine learning—are inherently nonconvex. Remarkably, a large class of such problems remains amenable to efficient optimization due to additional structure that weakens or generalizes convexity without forfeiting favorable algorithmic behavior. This paper surveys and systematizes notions of convexity and its generalizations, while also providing new comparative insights and explicit inclusion relationships among these function classes. We present a coherent taxonomy of functions that generalize, strengthen, and relax convexity, consolidating definitions, equivalent characterizations, closure properties, and hierarchical relations that are currently scattered across the optimization, operations research, and machine learning literature. Particular emphasis is placed on quasar-convexity, a recently introduced geometric condition that captures structured nonconvexity while enabling convergence guarantees comparable to those of convex optimization for many first-order methods. Through explicit inclusion diagrams and systematic comparisons, we clarify the relationships among classical generalizations, geometric variants, regularity conditions, and partial convexity notions. The resulting "convexity zoo" provides a comprehensive reference for researchers seeking to understand and exploit structured nonconvexity in contemporary optimization.

KEYWORDS
Nonconvex optimization; generalized convexity; quasar-convexity; star-convexity; structured nonconvexity; Polyak–Łojasiewicz inequality

1. Introduction

Convexity stands as arguably the most consequential structural property in mathematical optimization: when both the objective function and feasible region are convex, every local minimizer is automatically global, first-order optimality conditions are both necessary and sufficient, and a rich algorithmic toolkit—from gradient descent to interior-point methods—delivers polynomial-time convergence guarantees [20, 68]. This elegant framework has driven the remarkable success of convex optimization across signal processing, machine learning, statistics, and control. Yet the scope of practical optimization extends far beyond convexity; modern applications in reinforce-

CONTACT A. Khademi Email: abbaskhademi92@gmail.com

ment learning [31], deep learning [26], matrix factorization [25], and robust statistics [60] routinely involve objectives that are manifestly nonconvex. A central challenge in contemporary optimization research is therefore to identify and exploit *structure within nonconvexity*—properties weaker than convexity that nonetheless enable efficient algorithms with rigorous guarantees.

This challenge has spawned a rich and increasingly fragmented landscape of function classes that generalize, strengthen, or relax convexity in distinct directions. Classical generalizations such as quasi-convexity and pseudo-convexity, introduced in the 1960s [9, 62], preserve convex sublevel sets or the equivalence between stationarity and global optimality. Weakly convex (also termed para-convex or semi-convex) functions [24, 30, 80, 88] permit controlled deviation from convexity via quadratic majorization, thereby enabling proximal-type methods in nonsmooth nonconvex settings [27]. Regularity conditions—including the Polyak–Łojasiewicz (PL) inequality [76] and error bounds [61]—encode gradient dominance and ensure linear convergence without any convexity assumption. More recently, *geometric* generalizations have emerged as particularly powerful: star-convex functions, formalized in the context of cubic regularization [69], require the convexity inequality to hold only along rays emanating from a global minimizer, thereby accommodating local nonconvexity while preserving favorable global structure.

Among these, *quasar-convexity* has rapidly become a focal point for the algorithmic analysis of structured nonconvex problems. Initially introduced as "weak quasi-convexity" in [39] for linear dynamical system identification—and subsequently renamed in [42] to avoid confusion with classical quasi-convexity—quasar-convexity is parameterized by $\gamma \in (0, 1]$, continuously interpolating between star-convexity ($\gamma = 1$) and progressively weaker conditions as $\gamma \to 0^+$. This parametric flexibility captures a form of "structured nonconvexity" that is strict enough to enable fast convergence yet broad enough to encompass objectives arising in recurrent neural network training [37, 39], policy optimization [31], low-rank matrix recovery and phase retrieval [25], and modern neural loss landscapes [90]. Algorithmically, quasar-convex functions admit convergence rates closely mirroring those for convex optimization—representing a dramatic improvement over generic guarantees for smooth nonconvex functions and identifying a practically relevant "middle ground" between convexity and full nonconvexity.

Despite its growing importance, the theory surrounding quasar-convexity—and, more broadly, the entire ecosystem of convexity generalizations—remains scattered across disparate literatures. Definitions vary across communities (e.g., "weakly convex" in optimization vs. "para-convex" in functional analysis [82] or "semi-convex" in PDE theory [24]); implications are often stated without proof or under incompatible assumptions; and the mapping from structural properties to algorithmic consequences is frequently obscured by notational inconsistencies. This fragmentation impedes both research and practice: algorithm designers may be unaware of the precise function classes their methods exploit, practitioners lack a systematic framework for diagnosing problem structure, and newcomers face a bewildering array of disconnected definitions.

To address this, we present the first comprehensive, unified treatment of function classes enabling tractable nonconvex optimization. We provide a coherent taxonomy organized by structural theme—relaxations of convexity (quasi-convexity, pseudo-convexity, invexity, weak convexity), geometric generalizations (star-convexity, quasar-convexity and its variants), regularity conditions (PL, error bounds, quadratic growth), partial convexity (biconvex, multiconvex), and special classes (log-convex, DC)—supplying precise definitions, equivalent characterizations, and closure properties in consistent notation. We establish and visualize the hierarchical relationships among these classes through explicit inclusion diagrams, clarifying which implications are strict, which hold only under additional hypotheses (e.g., smoothness, compactness), and which fail in general. In particular, we deliver the most complete exposition to date of quasar-convexity—including strong, projected, and tilted variants—and position it as a central organizing concept linking classical notions (star-convexity, PL condition) to modern algorithmic guarantees. While prior works offer deep but narrow treatments—e.g., [33] on generalized convexity in economics, [24] on semi-concave functions, [49] on regularity conditions, and [34] on biconvex optimization—our survey is distinguished by its breadth, its emphasis on modern geometric classes emerging from machine learning, its focus on inter-class relationships, and its sustained algorithmic perspective.

The remainder of the paper is structured as follows. Section 2 reviews the foundations of classical convexity. Section 3 surveys traditional generalized convexity. Section 4 reviews weak convexity. Section 5 develops geometric generalizations, with emphasis on quasar-convexity and its algorithmic implications. Section 6 explores regularity conditions and their interrelations. Section 7 discusses partial and pointwise convexity. Section 8 covers special classes, including log-convex and DC functions. Finally, Section 9 synthesizes the entire landscape in the "Convexity Zoo," before concluding remarks in Section 10.

1.1. Notation and conventions

We work in $\mathbb{R}^n$ equipped with the norm $\|x\| = \sqrt{x^\top x}$. The extended real line is $\overline{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$. We denote by $\mathbb{R}_{++}$ the set of strictly positive real numbers. For a function $f\colon \mathcal{X} \to \overline{\mathbb{R}}$, we denote its epigraph by $\operatorname{epi}(f) = \{(x,t) \mid x \in \mathcal{X},\, f(x) \le t\}$, and its α-sublevel set by $\operatorname{lev}_\alpha(f) = \{x \in \mathcal{X} \mid f(x) \le \alpha\}$. A function f is said to be continuously differentiable on a set $\mathcal{C}$ if there exists an open set $U \supseteq \mathcal{C}$ such that f is defined and continuously differentiable on U. For differentiable f, $\nabla f(x)$ is the gradient and $\nabla^2 f(x)$ the Hessian (when it exists). A differentiable function $f\colon \mathcal{X} \to \mathbb{R}$ is said to be *L-smooth* (or have *L-Lipschitz continuous gradient*) for some $L > 0$ if

$$\|\nabla f(x) - \nabla f(y)\| \le L\|x - y\| \quad \text{for all } x, y \in \mathcal{X}. \tag{1}$$

For a symmetric matrix $A \in \mathbb{R}^{n\times n}$, we write $A \succeq 0$ ($A \succ 0$) to indicate that A is positive semidefinite (definite).

Throughout this paper, we consider optimization problems of the form

$$\min_{x\in\mathcal{X}} f(x), \tag{2}$$

where $\mathcal{X} \subseteq \mathbb{R}^n$ is a (typically convex) feasible set and $f : \mathcal{X} \to \overline{\mathbb{R}}$ is the objective function. The set of global minimizers is denoted $\mathcal{X}^\star = \operatorname{argmin}_{x\in\mathcal{X}} f(x)$, assumed nonempty where relevant. We write $f^\star := \inf_{x\in\mathcal{X}} f(x)$ for the optimal value (which equals $\min_{x\in\mathcal{X}} f(x)$ when the minimum is attained). If $\mathcal{X}^\star \neq \emptyset$, we let $x^\star \in \mathcal{X}^\star$ denote an arbitrary global minimizer, so that $f(x^\star) = f^\star$. We write $x^p := \operatorname{Proj}_{\mathcal{X}^\star}(x)$ for the Euclidean projection of x onto the set of global minimizers $\mathcal{X}^\star$.

2. Foundations of Convexity

Classical convexity theory forms the bedrock of modern optimization, with roots tracing back to Jensen's definition of convex functions in 1906 [45]. Valued for their structural simplicity and proximity to linearity among nonlinear functions [19], convex functions have become indispensable across optimization, economics, and operations research. In this section, we provide a systematic exposition of convex and strongly convex functions, focusing on the core characterizations and properties essential for understanding their generalizations in subsequent sections. Additional technical details and supplementary results are provided in Appendix A.

From this foundation, we formalize the basic geometric and analytic primitives upon which the theory rests.

2.1. Basic definitions

We begin with the foundational definitions of convex sets and functions.

Definition 2.1 (Convex Set)**.** A set $\mathcal{X} \subseteq \mathbb{R}^n$ is *convex* if for all $x, y \in \mathcal{X}$ and all $\lambda \in [0, 1]$,

$$\lambda x + (1 - \lambda) y \in \mathcal{X}.$$

□

Definition 2.2 (Convex Function)**.** A function $f : \mathcal{X} \to \mathbb{R}$ is *convex* if $\mathcal{X}$ is convex and, for all $x, y \in \mathcal{X}$ and $\lambda \in [0, 1]$,

$$f(\lambda x + (1 - \lambda) y) \leq \lambda f(x) + (1 - \lambda) f(y).$$

□

If the above inequality is strict whenever $x \neq y$ and $0 < \lambda < 1$, then f is *strictly convex*. A function is *concave* (resp., *strictly concave*) if $-f$ is convex (resp., strictly convex).

Remark 1. For extended-real-valued functions $f : \mathbb{R}^n \to \overline{\mathbb{R}}$, the convexity inequality in Definition 2.2 is required to hold only when $x, y \in \text{dom}(f)$. For the function to be convex, the effective domain $\text{dom}(f)$ must itself be convex. □

One of the beautiful aspects of convexity is the multiplicity of equivalent characterizations, each providing different geometric or analytic insights.

Property 2.3 (Jensen's Inequality)**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ be a convex set and $f : \mathcal{X} \to \mathbb{R}$. Then f is convex if and only if for every finite collection $x^1, \cdots, x^k \in \mathcal{X}$ and every set of weights $\lambda_1, \cdots, \lambda_k \geq 0$ satisfying $\sum_{i=1}^{k} \lambda_i = 1$,

$$f\left(\sum_{i=1}^{k} \lambda_i x^i\right) \leq \sum_{i=1}^{k} \lambda_i f(x^i).$$

□

Property 2.4 (Epigraph Criterion). Let $\mathcal{X} \subseteq \mathbb{R}^n$ be a convex set and $f : \mathcal{X} \to \mathbb{R}$. Then f is convex if and only if its epigraph, $\mathrm{epi}(f)$ is a convex subset of $\mathbb{R}^{n+1}$. $\square$

Property 2.5 (Sublevel Set Convexity). Let $f : \mathcal{X} \to \mathbb{R}$ be a convex function, where $\mathcal{X} \subseteq \mathbb{R}^n$ is convex. Then for every $\alpha \in \mathbb{R}$, the sublevel set $\mathrm{lev}_\alpha(f)$ is convex. $\square$

Remark 2. The converse of Property 2.5 does not hold: functions with convex sublevel sets are called *quasi-convex* (see Section 3) and need not be convex. For example, $f(x) = x^3$ on $\mathbb{R}$ has convex sublevel sets but is not convex.

When f is differentiable, convexity can be characterized through gradient inequalities and Hessian conditions.

Property 2.6 (Gradient Inequality). Let $f : \mathcal{X} \to \mathbb{R}$ be differentiable on a nonempty convex set $\mathcal{X} \subseteq \mathbb{R}^n$. Then f is convex on $\mathcal{X}$ if and only if for all $x, y \in \mathcal{X}$

$$f(y) \geq f(x) + \nabla f(x)^\top (y - x).$$

Moreover, if the inequality is strict for all $x \neq y$, then f is strictly convex. $\square$

The gradient inequality states that the first-order Taylor approximation globally underestimates a convex function.

Property 2.7 (Second-Order Characterization). Let $f : \mathcal{X} \to \mathbb{R}$ be twice continuously differentiable on a nonempty convex set $\mathcal{X} \subseteq \mathbb{R}^n$. Then:

(1) f is convex on $\mathcal{X}$ if and only if $\nabla^2 f(x) \succeq 0$ for all $x \in \mathcal{X}$.
(2) If $\nabla^2 f(x) \succ 0$ for all $x \in \mathcal{X}$, then f is strictly convex.

The converse of (2) is false: $f(x) = x^4$ is strictly convex on $\mathbb{R}$, but $f''(0) = 0$. $\square$

Additional equivalent characterizations, including the line restriction property, monotone secant slope criterion, and gradient monotonicity, are provided in Appendix A.1.

2.2. Subgradients and subdifferentials

Many convex functions arising in applications are not differentiable. The theory of subgradients extends differential calculus to this nonsmooth setting.

Definition 2.8 (Subgradient and Subdifferential). Let $f : \mathcal{X} \to \mathbb{R}$ be a convex function, where $\mathcal{X} \subseteq \mathbb{R}^n$ is convex. A vector $g \in \mathbb{R}^n$ is a *subgradient* of f at $x \in \mathcal{X}$ if for all $y \in \mathcal{X}$

$$f(y) \geq f(x) + g^\top (y - x).$$

$\square$

The set of all subgradients of f at x is the *subdifferential* of f at x, denoted by $\partial f(x)$. The subdifferential $\partial f(x)$ is a closed convex set; if $x \in \mathrm{int}(\mathcal{X})$, it is nonempty and bounded. When f is differentiable at x, we have $\partial f(x) = \{\nabla f(x)\}$.

2.3. Strengthenings of convexity

2.3.1. Strong convexity

Strong convexity strengthens classical convexity by imposing a uniform lower bound on curvature, leading to faster convergence rates in optimization algorithms.

Definition 2.9 (Strong Convexity)**.** Let $\mu > 0$ and let $\mathcal{X} \subseteq \mathbb{R}^n$ be a convex set. A function $f : \mathcal{X} \to \mathbb{R}$ is *μ-strongly convex* if for all $x, y \in \mathcal{X}$ and all $\lambda \in [0, 1]$,

$$f\big(\lambda x + (1-\lambda)y\big) \le \lambda f(x) + (1-\lambda)f(y) - \frac{\mu}{2}\lambda(1-\lambda)\|x-y\|^2.$$

□

Property 2.10 (Strong First-Order Condition)**.** Let $\mu > 0$. Let $f : \mathcal{X} \to \mathbb{R}$ be continuously differentiable on a convex set $\mathcal{X} \subseteq \mathbb{R}^n$. Then f is μ-strongly convex if and only if for all $x, y \in \mathcal{X}$

$$f(y) \ge f(x) + \nabla f(x)^\top (y-x) + \frac{\mu}{2}\|y-x\|^2.$$

□

Property 2.11 (Hessian Lower Bound)**.** Let $f : \mathcal{X} \to \mathbb{R}$ be twice continuously differentiable on a convex set $\mathcal{X} \subseteq \mathbb{R}^n$. Then f is μ-strongly convex if and only if for all $x \in \mathcal{X}$, $\nabla^2 f(x) \succeq \mu I$. □

Property 2.12 (Quadratic Perturbation Characterization)**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ be a convex set and $\mu > 0$. A function $f : \mathcal{X} \to \mathbb{R}$ is μ-strongly convex on $\mathcal{X}$ if and only if the function $f(x) - \frac{\mu}{2}\|x\|^2$ is convex on $\mathcal{X}$. □

Additional characterizations and properties of strong convexity are provided in Appendix A.3.

2.3.2. Uniform convexity

Uniform convexity generalizes strong convexity by replacing the quadratic modulus with more general growth conditions [13, 91].

Definition 2.13 (Uniformly Convex Function)**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ be a convex set. A function $f : \mathcal{X} \to \mathbb{R}$ is *uniformly convex* with modulus $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ if ϕ is nondecreasing, $\phi(0) = 0$, $\phi(t) > 0$ for all $t > 0$, and for all $x, y \in \mathcal{X}$ and all $\lambda \in [0, 1]$,

$$f\big(\lambda x + (1-\lambda)y\big) \le \lambda f(x) + (1-\lambda)f(y) - \lambda(1-\lambda)\,\phi\big(\|x-y\|\big).$$

□

When $\phi(t) = \frac{\mu}{2}t^2$, we recover μ-strong convexity. The case $\phi(t) = ct^p$ for $p > 1$ yields *p-uniformly convex* functions.

The various notions of convexity form a strict hierarchy:

$$\text{Strongly convex} \implies \text{Uniformly convex} \implies \text{Strictly convex} \implies \text{Convex}$$

2.4. Key optimality and closure properties

Convexity dramatically simplifies optimization by ensuring local optimality implies global optimality.

Property 2.14 (Local Equals Global)**.** Let $f : \mathcal{X} \to \mathbb{R}$ be a convex function on a convex set $\mathcal{X} \subseteq \mathbb{R}^n$. Then every local minimizer of f over $\mathcal{X}$ is a global minimizer. □

Property 2.15 (First-Order Optimality)**.** Let $f : \mathcal{X} \to \mathbb{R}$ be convex and differentiable on a nonempty convex set $\mathcal{X} \subseteq \mathbb{R}^n$. Then $x^\star \in \mathcal{X}$ is a global minimizer of f over $\mathcal{X}$ if and only if for all $x \in \mathcal{X}$

$$\nabla f(x^\star)^\top (x - x^\star) \geq 0.$$

In the unconstrained case ($\mathcal{X} = \mathbb{R}^n$), this reduces to $\nabla f(x^\star) = 0$. □

Convex functions are closed under numerous operations including nonnegative weighted sums, affine precomposition, composition with nondecreasing convex functions, and pointwise suprema. These closure properties and additional optimality conditions are detailed in Appendix A.5.

Collectively, these properties and characterizations constitute the core toolkit for classical convex analysis. In the following sections, we systematically relax or replace these assumptions—first with quasi-convexity and generalized monotonicity, then with structured nonconvex models—to build a taxonomy of function classes capable of capturing richer, real-world phenomena while preserving algorithmic tractability where possible.

3. Traditional generalized convexity

Generalized convexity plays a fundamental role in optimization, particularly in modeling nonconvex problems that retain sufficient geometric structure for tractable analysis and algorithms. This section presents the principal generalizations of convexity—quasi-convexity, pseudo-convexity, invexity, and r-convexity—which preserve key optimization properties while accommodating broader function classes [23, 33]. Additional variants and technical characterizations are provided in Appendix B.

3.1. Quasi-convexity

Quasi-convex functions were among the first and most influential generalizations of convexity.

Definition 3.1 (Quasi-convex Function)**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ be a convex set and $f : \mathcal{X} \to \mathbb{R}$. The function f is *quasi-convex* on $\mathcal{X}$ if for all $x, y \in \mathcal{X}$ and all $\lambda \in [0, 1]$,

$$f\big(\lambda x + (1 - \lambda)y\big) \leq \max\big\{f(x),\ f(y)\big\}.$$

□

Property 3.2 (Sublevel Set Characterization)**.** A function $f : \mathcal{X} \to \mathbb{R}$ is quasi-convex if and only if, for every $\alpha \in \mathbb{R}$, the sublevel set $\mathrm{lev}_\alpha(f)$ is convex (with the convention that $\emptyset$ is convex). □

Definition 3.3 (Strictly Quasi-convex Function)**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ be convex and $f : \mathcal{X} \to \mathbb{R}$. The function f is *strictly quasi-convex* on $\mathcal{X}$ if for all $x, y \in \mathcal{X}$ with $x \neq y$ and all $\lambda \in (0, 1)$,

$$f\big((1-\lambda)x + \lambda y\big) < \max\big\{f(x),\, f(y)\big\}.$$

□

Definition 3.4 (Semistrictly Quasi-convex Function)**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ be a convex set and $f : \mathcal{X} \to \mathbb{R}$. The function f is *semistrictly quasi-convex* on $\mathcal{X}$ if it is quasi-convex and if for all $x, y \in \mathcal{X}$ with $f(x) < f(y)$ and for all $\lambda \in (0, 1)$,

$$f\big(\lambda x + (1-\lambda)y\big) < f(y).$$

□

Property 3.5 (First-Order Characterization of Quasi-convexity)**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ be convex and $f : \mathcal{X} \to \mathbb{R}$ be differentiable on $\mathcal{X}$. Then f is quasi-convex on $\mathcal{X}$ if and only if for all $x, y \in \mathcal{X}$ with $f(x) \leq f(y)$

$$\nabla f(y)^\top (x - y) \leq 0.$$

Equivalently, for every $y \in \mathcal{X}$, the half-space $\big\{x \in \mathbb{R}^n \mid \nabla f(y)^\top (x - y) \leq 0\big\}$ contains the sublevel set $\{x \in \mathcal{X} \mid f(x) \leq f(y)\}$. □

3.2. Pseudo-convexity

A defining property of differentiable convex functions is that every stationary point is a global minimizer. This property, however, is not exclusive to convexity. The family of pseudo-convex functions, introduced in [62], strictly includes the family of differentiable convex functions and has the above-mentioned property.

Definition 3.6 (Pseudo-convex Function)**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ be a convex set and let $f : \mathcal{X} \to \mathbb{R}$ be differentiable on $\mathcal{X}$. The function f is *pseudo-convex* on $\mathcal{X}$ if for all $x, y \in \mathcal{X}$,

$$f(x) < f(y) \Longrightarrow \nabla f(y)^\top (x - y) < 0.$$

Equivalently, f is pseudo-convex on $\mathcal{X}$ if for all $x, y \in \mathcal{X}$,

$$\nabla f(y)^\top (x - y) \geq 0 \Longrightarrow f(x) \geq f(y).$$

□

From Definition 3.6, it follows immediately that if f is pseudo-convex on $\mathcal{X}$ and $\nabla f(\bar{x}) = 0$ for some $\bar{x} \in \mathcal{X}$, then $\bar{x}$ is a global minimizer of f over $\mathcal{X}$. Consequently, pseudo-convexity plays a pivotal role in nonlinear programming: when the objective function is differentiable and pseudo-convex (and the feasible region is convex), first-order stationarity is not merely necessary but also *sufficient* for global optimality.

Definition 3.7 (Strictly Pseudo-convex Function)**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ be a convex set and let $f : \mathcal{X} \to \mathbb{R}$ be differentiable on $\mathcal{X}$. The function f is *strictly pseudo-convex* on $\mathcal{X}$ if for all $x, y \in \mathcal{X}$ with $x \neq y$,

$$f(x) \leq f(y) \Longrightarrow \nabla f(y)^\top (x - y) < 0.$$

Equivalently, f is strictly pseudo-convex on $\mathcal{X}$ if for all $x, y \in \mathcal{X}$ with $x \neq y$,

$$\nabla f(y)^\top (x - y) \geq 0 \Longrightarrow f(x) > f(y).$$

□

3.3. Invexity

Invex functions, introduced in [38], generalize classical convexity by allowing a flexible "direction" mapping η in place of the standard displacement $x - y$. In recent years, invexity and its generalizations have emerged as relevant structural assumptions in signal processing and machine learning [12, 71], particularly in the analysis of nonconvex models where classical convexity is violated but first-order optimality conditions remain sufficient for global optimality [15].

Definition 3.8 (Invex Function)**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ be a nonempty set and let $f : \mathcal{X} \to \mathbb{R}$ be differentiable on $\mathcal{X}$. The function f is *invex* on $\mathcal{X}$ if there exists a vector-valued mapping $\eta : \mathcal{X} \times \mathcal{X} \to \mathbb{R}^n$ such that for all $x, y \in \mathcal{X}$,

$$f(x) - f(y) \geq \eta(x, y)^\top \nabla f(y).$$

□

Property 3.9 (Stationary-Point Characterization of Invexity)**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ be nonempty and let $f \colon \mathcal{X} \to \mathbb{R}$ be differentiable on $\mathcal{X}$. Then f is invex on $\mathcal{X}$ *with respect to some mapping* η if and only if every stationary point is a global minimizer; that is, for all $x^0 \in \mathcal{X}$,

$$\nabla f(x^0) = 0 \Longrightarrow f(x^0) = \inf_{x \in \mathcal{X}} f(x).$$

□

The following properties establish the relationships among these function classes.

Property 3.10 (Convexity Implies Pseudo-convexity)**.** Every convex differentiable function is pseudo-convex. The converse is false. □

Property 3.11 (Pseudo-convexity Implies Quasi-convexity)**.** Every pseudo-convex function is quasi-convex. The converse is false. □

Property 3.12 (Convexity Implies Invexity)**.** Let $f \colon \mathcal{X} \to \mathbb{R}$ be differentiable and convex on $\mathcal{X}$. Then f is invex on $\mathcal{X}$ with respect to the mapping $\eta(x, y) = x - y$. □

Thus, for differentiable functions on convex domains, we have the hierarchy:

$$\text{Convex} \Longrightarrow \text{Pseudo-convex} \Longrightarrow \text{Quasi-convex}$$

and separately:

$$\text{Convex} \implies \text{Invex}$$

Definition 3.13 (Quasi-Invex Function)**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ be a convex set and let $f : \mathcal{X} \to \mathbb{R}$ be differentiable on $\mathcal{X}$. The function f is *quasi-invex* on $\mathcal{X}$ if there exists a vector-valued mapping $\eta : \mathcal{X} \times \mathcal{X} \to \mathbb{R}^n$, not identically zero, such that for all $x, y \in \mathcal{X}$,

$$\eta(x, y)^\top \nabla f(y) > 0 \implies f(x) > f(y).$$

□

Property 3.14. A differentiable quasi-convex function is also quasi-invex □

Property 3.15 (Invex Implies Quasi-Invex)**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ be convex, and let $f : \mathcal{X} \to \mathbb{R}$ be differentiable. If f is invex on $\mathcal{X}$ with mapping η, then f is quasi-invex on $\mathcal{X}$ with the same kernel function η. The converse does not hold in general. □

For proofs and additional details, see, e.g., [33]. Further variants of quasi-convexity (strong, uniform, neat, explicit) and second-order characterizations are provided in Appendix B. We also note that invex functions and related concepts have recently attracted considerable attention in signal processing and machine learning; see [71] and the references therein.

3.4. r-convexity

This subsection introduces the notion of r-convexity, a parametric generalization of convexity that is naturally expressed via power means and exponential transformations. The concept interpolates between quasi-concavity, classical convexity, and quasi-convexity through limiting values of the parameter r. We follow the standard definitions and characterizations in [11, 94].

Definition 3.16 (Generalized r-th Mean)**.** Let $\alpha, \beta > 0$ and let $\lambda \in [0, 1]$. The *generalized r-th mean* (or *power mean of order r*) of α and β with weight λ is defined by

$$M_r(\alpha, \beta; \lambda) := \begin{cases} \big(\lambda\alpha^r + (1-\lambda)\beta^r\big)^{1/r}, & r \in \mathbb{R} \setminus \{0\}, \\ \lim_{r\to 0} M_r(\alpha, \beta; \lambda) = \alpha^\lambda \beta^{1-\lambda}, & r = 0, \\ \lim_{r\to +\infty} M_r(\alpha, \beta; \lambda) = \max\{\alpha, \beta\}, & r = +\infty, \\ \lim_{r\to -\infty} M_r(\alpha, \beta; \lambda) = \min\{\alpha, \beta\}, & r = -\infty. \end{cases}$$

□

Definition 3.17 (r-Convex Function)**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ be a convex set and let $r \in \overline{\mathbb{R}}$. A function $f : \mathcal{X} \to \mathbb{R}$ is *r-convex* on $\mathcal{X}$ if for all $x, y \in \mathcal{X}$ and all $\lambda \in [0, 1]$,

$$f\big(\lambda x + (1-\lambda)y\big) \le \log M_r\big(e^{f(x)}, e^{f(y)}; \lambda\big).$$

□

Property 3.18 (Interpretation of Limiting Cases)**.** The limiting cases correspond to classical function classes:

(1) **Case** $r = 0$**:** The function f is *convex* in the classical sense (Definition 2.2).
(2) **Case** $r = +\infty$**:** The function f is *quasi-convex* (Definition 3.1).
(3) **Case** $r = -\infty$**:** The function f is *quasi-concave*, i.e., $-f$ is quasi-convex.

Thus, r-convexity provides a unified parametric framework interpolating between quasi-concavity, convexity, and quasi-convexity. □

Property 3.19 (Exponential Transformation Characterization)**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ be a convex set and let $r \in \mathbb{R} \setminus \{0\}$. A function $f : \mathcal{X} \to \mathbb{R}$ is r-convex on $\mathcal{X}$ if and only if the transformed function $\hat{f}(x) := e^{rf(x)}$ satisfies:

(1) $\hat{f}$ is convex on $\mathcal{X}$ if $r > 0$;
(2) $\hat{f}$ is concave on $\mathcal{X}$ if $r < 0$.

□

Property 3.20 (Second-Order Characterization of r-Convexity)**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ be a convex set and let $f : \mathcal{X} \to \mathbb{R}$ be twice continuously differentiable. Then f is r-convex on $\mathcal{X}$ if and only if the matrix

$$Q(x) := r\,\nabla f(x)\nabla f(x)^\top + \nabla^2 f(x)$$

is positive semidefinite for all $x \in \mathcal{X}$, i.e.,

$$Q(x) = r\,\nabla f(x)\nabla f(x)^\top + \nabla^2 f(x) \succeq 0, \quad \forall\, x \in \mathcal{X}.$$

□

Property 3.21 (Optimal r-Convexity Parameter for Quasi-Convex Functions)**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ be an open convex set and let $f : \mathcal{X} \to \mathbb{R}$ be twice continuously differentiable and quasi-convex on $\mathcal{X}$. Define

$$r^\star := \sup_{\substack{x \in \mathcal{X},\, z \in \mathbb{R}^n \\ \|z\|=1,\, z^\top \nabla f(x) \neq 0}} \frac{-z^\top \nabla^2 f(x)\, z}{\big(z^\top \nabla f(x)\big)^2}.$$

If $r^\star < +\infty$, then f is $r^\star$-convex on $\mathcal{X}$. Moreover, $r^\star$ is the smallest value of r for which f is r-convex. □

4. Weak Convexity and Its Variants

Weakly convex functions—also known as *para-convex* in functional analysis [80–82] and *semi-convex* in nonsmooth analysis and PDE theory [24]—constitute a principled relaxation of convexity that preserves much of its algorithmic power [48, 81]. In PDE theory, semi-convexity and semi-concavity are often defined up to a sign convention; both correspond to weak convexity of either f or $-f$.

This function class arises naturally in nonsmooth optimization, phase retrieval, neural network training, and Moreau envelope smoothing, where nonconvex objectives exhibit controlled deviations from convexity. A defining feature of weak convexity is that adding a suitable quadratic term restores convexity, thereby enabling the extension of proximal and subgradient methods to nonconvex settings with rigorous convergence guarantees [27, 30].

Definition 4.1 (Weakly Convex (Para-convex) Function). Let $\mathcal{X} \subseteq \mathbb{R}^n$ be a convex set. A function $f : \mathcal{X} \to \mathbb{R}$ is *ρ-weakly convex* (or *ρ-para-convex*) on $\mathcal{X}$ with modulus $\rho \geq 0$ if for all $x, y \in \mathcal{X}$ and all $\lambda \in [0, 1]$,

$$f\big(\lambda x + (1-\lambda)y\big) \leq \lambda f(x) + (1-\lambda)f(y) + \frac{\rho}{2}\lambda(1-\lambda)\,\|x-y\|^2.$$

□

The following characterization is the cornerstone of algorithmic extensions: minimizing f is equivalent to minimizing a convex function after an additive quadratic transformation.

Property 4.2 (Quadratic Convexification). Let $\mathcal{X} \subseteq \mathbb{R}^n$ be convex and $f : \mathcal{X} \to \mathbb{R}$. Then f is ρ-weakly convex if and only if the function $f(x) + \frac{\rho}{2}\|x\|^2$ is convex on $\mathcal{X}$. □

Property 4.3 (First-Order Characterization). Let $f : \mathcal{X} \to \mathbb{R}$ be differentiable on a convex set $\mathcal{X} \subseteq \mathbb{R}^n$. Then f is ρ-weakly convex if and only if for all $x, y \in \mathcal{X}$,

$$f(y) \geq f(x) + \nabla f(x)^\top (y-x) - \frac{\rho}{2}\|y-x\|^2.$$

□

This inequality states that the first-order Taylor model, corrected by a quadratic term that controls the allowable nonconvexity, globally lower-bounds f. It generalizes the gradient inequality for convex functions (Property 2.6) and is instrumental in convergence analysis of first-order methods.

Property 4.4 (Second-Order Characterization). Let $f : \mathcal{X} \to \mathbb{R}$ be twice continuously differentiable on a convex set $\mathcal{X} \subseteq \mathbb{R}^n$. Then f is ρ-weakly convex if and only if for all $x \in \mathcal{X}$

$$\nabla^2 f(x) \succeq -\rho I,$$

i.e., all eigenvalues of $\nabla^2 f(x)$ are bounded below by $-\rho$. Compare with the strong convexity condition in Property 2.11. Weak convexity allows indefinite Hessians, provided negative curvature is uniformly bounded. □

Property 4.5 (Smoothness Implies Weak Convexity). Let $f : \mathcal{X} \to \mathbb{R}$ be differentiable on a convex set $\mathcal{X} \subseteq \mathbb{R}^n$. If ∇f is L-smooth, then:

(1) f is L-weakly convex;
(2) $-f$ is L-weakly convex.

□

Thus, any L-smooth function and its negative are both L-weakly convex—a fact important in minimax and bilevel optimization.

Remark 3 (Terminology)**.** The terms *weakly convex* [88], *para-convex* [82], and *semi-convex* [24] are used interchangeably in modern literature, with minor variations in constant conventions. All refer to the same structural property: convexity up to a quadratic penalty. □

4.1. Generalized para-convexity

The notion of weak convexity can be extended by allowing subquadratic or superquadratic deviation from convexity. Introduced in [80] and further developed in [48, 70], these classes model functions whose nonconvexity is controlled by power-type error terms. They arise in Hölder-smooth optimization, robust estimation, and regularized inverse problems where quadratic bounds are too restrictive.

Definition 4.6 ((ν,ρ)-Para-convex Function)**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ be a convex set, let $\nu \in (0,1]$, and let $\rho \geq 0$. A function $f : \mathcal{X} \to \mathbb{R}$ is (ν,ρ)-*para-convex* on $\mathcal{X}$ with modulus ρ if for all $x, y \in \mathcal{X}$ and all $\lambda \in [0,1]$,

$$f\big(\lambda x + (1-\lambda)y\big) \leq \lambda f(x) + (1-\lambda)f(y) + \frac{\rho}{2}\min\{\lambda, 1-\lambda\}\,\|x-y\|^{1+\nu}.$$

□

The exponent $1+\nu$ captures subquadratic ($\nu < 1$) or quadratic ($\nu = 1$) deviation from convexity.

Property 4.7 (Recovery of Weak Convexity)**.** Let $f : \mathcal{X} \to \mathbb{R}$ be continuous on a convex set $\mathcal{X} \subseteq \mathbb{R}^n$. Then f is $(1,\rho)$-para-convex if and only if f is weakly convex in the sense of Definition 4.1. □

Thus, ν-para-convexity with $\nu < 1$ strictly generalizes weak convexity, enabling analysis of functions with slower-than-quadratic curvature decay (e.g., $f(x) = \|x\|^{1+\nu}$ near zero).

The structural richness of weakly convex functions is underscored by their role in difference-of-convex (DC) programming and as building blocks for advanced optimization methods. Recent work investigates their Moreau envelope properties [79], approximate subdifferentials [87], inexact proximal schemes [53], and applications to robust low-rank recovery [55, 77]. Extensions to composite and stochastic settings, as well as learning-theoretic perspectives [10, 17, 35], attest to the continued relevance of this function class.

5. Modern Geometric Generalizations of Convexity

This section presents two geometrically motivated generalizations of convexity—*star-convexity* and *quasar-convexity*—that have emerged as particularly relevant for modern optimization, especially in machine learning. These classes retain sufficient structure to support efficient first-order methods with provable convergence guarantees, while accommodating objectives that are inherently nonconvex yet structured.

The key insight underlying both notions is that global optimization does not require convexity along *all* line segments. Instead, it suffices to impose favorable geometry along paths leading toward global minimizers. This perspective leads to one-sided conditions relative to optimal points, rather than the symmetric conditions characteristic of classical convexity.

5.1. Star-convexity

Star-convex functions constitute an important class of nonconvex functions that properly contains all convex functions [69].

5.1.1. Basic definitions and geometric intuition

Definition 5.1 (Star-Shaped Set)**.** A set $S \subseteq \mathbb{R}^n$ is *star-shaped with respect to a point* $x^\star \in S$ if for every $x \in S$ and every $\lambda \in [0,1]$,

$$\lambda x^\star + (1-\lambda)x \in S.$$

The point $x^\star$ is called a *star center* of S. Equivalently, S is star-shaped at $x^\star$ if the line segment $[x^\star, x] \subseteq S$ for all $x \in S$. □

Every convex set is star-shaped with respect to each of its points. However, a star-shaped set need not be convex. The notion of star-convexity extends this geometric concept to functions.

Definition 5.2 (Star-Convex Function)**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ be a convex set and let $f : \mathcal{X} \to \mathbb{R}$. The function f is *star-convex* on $\mathcal{X}$ if the set of global minimizers $\mathcal{X}^\star$ is nonempty, and for every $x^\star \in \mathcal{X}^\star$, every $x \in \mathcal{X}$, and every $\lambda \in [0,1]$,

$$f\big(\lambda x^\star + (1-\lambda)x\big) \le \lambda f(x^\star) + (1-\lambda) f(x).$$

□

This definition requires the star-convexity property to hold uniformly for all global minimizers [69]. Some works consider a weaker, center-specific notion where the inequality need only hold for a single global minimizer [42, 57].

Definition 5.3 (Star-Convex with Respect to a Point)**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ be a convex set, let $x^\star \in \mathcal{X}$, and let $f : \mathcal{X} \to \mathbb{R}$. The function f is *star-convex at* $x^\star$ on $\mathcal{X}$ if for all $x \in \mathcal{X}$ and all $\lambda \in [0,1]$,

$$f\big(\lambda x^\star + (1-\lambda)x\big) \le \lambda f(x^\star) + (1-\lambda) f(x).$$

In particular, if $x^\star \in \mathcal{X}^\star$, then f is said to be star-convex at a global minimizer. □

More generally, f is *strictly star-convex with respect to a point* $x^\star \in \mathcal{X}$ if the same strict inequality holds for all $x \in \mathcal{X} \setminus \{x^\star\}$ and all $\lambda \in (0,1)$.

Intuitively, when visualizing the objective function as a landscape, star-convexity ensures that each global optimum is "visible" from every feasible point. More precisely, along any ray from a point x toward any global minimizer $x^\star$, the function values decrease monotonically (in a weighted sense). This visibility property implies there are no ridges obstructing direct paths to global optima, though ridges may exist in orthogonal directions. This geometric structure suggests that gradient-based descent methods should be effective for this function class, as following the negative gradient will generally make progress toward the optimum.

Property 5.4 (Convexity Implies Star-Convexity)**.** Let $f : \mathcal{X} \to \mathbb{R}$ be convex on a convex set $\mathcal{X} \subseteq \mathbb{R}^n$, and suppose the set of global minimizers $\mathcal{X}^\star$ is nonempty. Then f is star-convex with respect to every $x^\star \in \mathcal{X}^\star$—in particular, f is star-convex. Moreover, if f is strictly convex, then it is strictly star-convex (and $\mathcal{X}^\star$ is a singleton). □

The converse does not hold: star-convexity is strictly weaker than convexity. This gap is precisely what makes star-convexity valuable—it captures a broader class of functions while retaining key optimization-friendly properties.

Definition 5.5 (Strong Star-Convexity)**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ be a convex set, let $f : \mathcal{X} \to \mathbb{R}$, and let $\mu > 0$. Suppose that the set of global minimizers $\mathcal{X}^\star$ is nonempty. The function f is *μ-strongly star-convex with respect to* $x^\star \in \mathcal{X}^\star$ if for all $x \in \mathcal{X}$ and all $\lambda \in [0, 1]$,

$$f\big(\lambda x^\star + (1 - \lambda)x\big) \leq \lambda f(x^\star) + (1 - \lambda)f(x) - \frac{\mu}{2}\,\lambda(1 - \lambda)\,\|x^\star - x\|^2.$$

□

Strong star-convexity, in the context of cubic-regularized Newton methods, strengthens star-convexity by adding a quadratic margin [69]. This additional structure ensures linear convergence for both gradient and proximal point methods [28, 42].

Property 5.6 (Strong Convexity Implies Strong Star-Convexity)**.** Strong convexity is a special case of strong star-convexity. Specifically, if f is μ-strongly convex on $\mathcal{X}$, then f is μ-strongly star-convex with respect to its unique minimizer $x^\star$. □

5.1.2. Characterizations of star-convexity

Star-convexity admits multiple equivalent characterizations that provide different perspectives on this function class.

Property 5.7 (Epigraph Characterization)**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ be convex, and let $f : \mathcal{X} \to \mathbb{R}$ be a function with nonempty set of global minimizers $\mathcal{X}^\star$. For any $x^\star \in \mathcal{X}^\star$, the function f is star-convex at $x^\star$ if and only if its epigraph, $\mathrm{epi}(f)$ is star-shaped with respect to the point $(x^\star, f(x^\star))$. □

This characterization provides a direct geometric interpretation: star-convexity of f corresponds to star-shapedness of its epigraph, just as convexity of f corresponds to convexity of its epigraph.

Property 5.8 (Star-Shaped Sublevel Sets)**.** Let $f : \mathcal{X} \to \mathbb{R}$ be star-convex with nonempty set of global minimizers $\mathcal{X}^\star$. Then for every $\alpha \in \mathbb{R}$, the sublevel set $\mathrm{lev}_\alpha(f)$ is star-shaped with respect to every $x^\star \in \mathcal{X}^\star$. □

Property 5.9 (First-Order Characterization of Star-Convexity)**.** Let $f : \mathcal{X} \to \mathbb{R}$ be differentiable on a convex set $\mathcal{X} \subseteq \mathbb{R}^n$, and assume the set of global minimizers $\mathcal{X}^\star$ is nonempty. If f is star-convex, then for every $x^\star \in \mathcal{X}^\star$ and every $x \in \mathcal{X}$,

$$f(x^\star) \geq f(x) + \nabla f(x)^\top (x^\star - x).$$

□

This first-order condition states that the tangent hyperplane at any point x lies below the optimal value $f(x^\star)$ when evaluated in the direction of $x^\star$. Compared to the gradient inequality for convex functions, this is a one-sided condition that need only hold in the direction toward the optimum.

Property 5.10 (First-Order Characterization of Strong Star-Convexity)**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ be convex, and let $f : \mathcal{X} \to \mathbb{R}$ be differentiable. Suppose the set of global minimizers $\mathcal{X}^\star$ is nonempty, and let $x^\star \in \mathcal{X}^\star$. Then f is μ-strongly star-convex at $x^\star$ if and only if for all $x \in \mathcal{X}$,

$$f(x^\star) \geq f(x) + \nabla f(x)^\top (x^\star - x) + \frac{\mu}{2}\|x - x^\star\|^2.$$

□

Property 5.11 (Necessary Second-Order Condition for Star-Convexity)**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ be open and star-shaped with respect to $x^0 \in \mathcal{X}$, and let $f : \mathcal{X} \to \mathbb{R}$ be twice continuously differentiable. If f is star-convex on $\mathcal{X}$ at $x^\star$, then the Hessian of f at $x^\star$ is positive semidefinite:

$$\nabla^2 f(x^\star) \succeq 0.$$

□

This necessary condition shows that star-convexity imposes local curvature requirements at the star center, though it does not require positive semidefiniteness of the Hessian at all points (as convexity does).

5.1.3. Algorithmic implications

The geometric structure of star-convex functions—particularly the visibility property—ensures that gradient-based methods can effectively locate global minimizers without getting trapped in spurious local optima. The concept was introduced in [69] to study cubic-regularized Newton methods for unconstrained optimization, and has since been extended to various algorithmic settings [28, 47, 52, 57].

For star-convex functions, gradient descent with appropriate step sizes converges to the global optimum at rate $O(1/N)$, matching the rate for convex functions. For strongly star-convex functions, linear convergence is achievable, paralleling the behavior of strongly convex optimization.

5.2. Quasar-convexity

The class of quasar-convex functions provides a principled framework for analyzing structured nonconvex optimization problems amenable to efficient first-order methods. Introduced initially as *weak quasi-convexity* in [39] and [37] in the context of linear dynamical system identification, the notion was later renamed *quasar-convexity* in [42] to avoid ambiguity with classical quasi-convexity. The name deliberately evokes a conceptual kinship with quasi-convexity while underscoring a distinct—indeed incomparable—functional geometry.

5.2.1. Basic definitions and motivation

The defining feature of quasar-convexity is a single parameter $\gamma \in (0, 1]$ that controls the degree of nonconvexity. When $\gamma = 1$, the class reduces to star-convexity, the classical geometric generalization [69]. As γ decreases from 1, progressively stronger deviations from convexity are permitted, yet the structure remains sufficiently regular to enable efficient first-order optimization methods with provable convergence guarantees. This "structured nonconvexity" has emerged as particularly relevant in modern machine learning applications. Training objectives for linear dynamical system identification exhibit quasar-convex structure [39], and certain overparameterized neural network architectures yield quasar-convex loss landscapes [90]. Policy optimization objectives in reinforcement learning and low-rank matrix recovery problems under appropriate conditions also fall within this framework.

From an algorithmic perspective, quasar-convex functions admit convergence rates that closely mirror those for convex optimization while encompassing a substantially broader class of objectives. Quasar-convexity thus captures a practically relevant "middle ground" between convexity and full nonconvexity.

Quasar-convexity admits a clean first-order characterization, which is often more amenable to algorithmic analysis.

Definition 5.12 (Quasar-Convexity)**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ be a convex set and let $f : \mathcal{X} \to \mathbb{R}$ be continuously differentiable on $\mathcal{X}$. Suppose that the set of global minimizers $\mathcal{X}^\star$ is nonempty, and fix $x^\star \in \mathcal{X}^\star$. For $\gamma \in (0, 1]$, the function f is *γ-quasar-convex on $\mathcal{X}$ at $x^\star$* if for all $x \in \mathcal{X}$,

$$f(x) - f(x^\star) \leq \frac{1}{\gamma} \nabla f(x)^\top (x - x^\star).$$

□

When $\gamma = 1$, the inequality reduces to the first-order condition for star-convexity at $x^\star$.

Remark 4 (Domain Assumptions)**.** The assumption that $\mathcal{X}$ be convex may be relaxed. Definition 5.12 remains well-posed under the weaker condition that $\mathcal{X}$ is *star-convex at $x^\star$*. In the unconstrained setting where $\mathcal{X} = \mathbb{R}^n$, this condition is trivially satisfied.

The following property establishes the fundamental relationship between quasar-convexity and star-convexity.

Property 5.13 (Star-Convexity as a Special Case)**.** Let $f : \mathcal{X} \to \mathbb{R}$ be continuously differentiable on a convex set $\mathcal{X} \subseteq \mathbb{R}^n$. If f is star-convex at $x^\star \in \mathcal{X}$, then f is 1-quasar-convex at $x^\star$. Conversely, if f is 1-quasar-convex at $x^\star$, then f is star-convex at $x^\star$. □

Since convexity implies star-convexity (Property 5.4), we have the chain of implications: convex functions are star-convex, which is equivalent to 1-quasar-convexity, which in turn implies γ-quasar-convexity for all $\gamma \in (0, 1]$.

Property 5.14 (Quasar-Convexity Parameter Ordering)**.** Let $f : \mathcal{X} \to \mathbb{R}$ be γ_1-quasar-convex at $x^\star$ for some $\gamma_1 \in (0, 1]$. Then f is γ_2-quasar-convex at $x^\star$ for all $\gamma_2 \in (0, \gamma_1]$. □

This property shows that larger values of γ correspond to stronger conditions: a function that is γ-quasar-convex for larger γ is "closer to convex" in a precise sense.

5.2.2. Optimality of stationary points

A defining characteristic of quasar-convex functions is that first-order stationary points are global minimizers—a hallmark of well-behaved nonconvex landscapes that enables the use of local search methods with global optimality guarantees.

Property 5.15 (First-Order Stationary Points Are Global Minimizers)**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ be a nonempty convex set and $f : \mathcal{X} \to \mathbb{R}$ be continuously differentiable. Suppose f is γ-quasar-convex on $\mathcal{X}$ at a global minimizer $x^\star \in \mathcal{X}^\star$ for some $\gamma \in (0, 1]$. Then any first-order stationary point $x \in \mathcal{X}$—i.e., a point satisfying $\nabla f(x)^\top (y - x) \geq 0$ for all $y \in \mathcal{X}$—is a global minimizer of f on $\mathcal{X}$, and satisfies $f(x) = f(x^\star)$. □

This property is fundamental for the algorithmic tractability of quasar-convex optimization: methods that converge to stationary points automatically find global optima, eliminating the problem of spurious local minima that plagues general nonconvex optimization.

Property 5.16 (Quasar-Convexity Implies Invexity)**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ be convex and $f : \mathcal{X} \to \mathbb{R}$ be continuously differentiable and γ-quasar-convex at $x^\star \in \mathcal{X}^\star$ for some $\gamma \in (0, 1]$. Then f is invex with respect to the mapping $\eta(x, y) = \frac{1}{\gamma}(x^\star - x)$. Consequently, every stationary point is a global minimizer. □

5.2.3. Strong quasar-convexity

Analogous to the relationship between convexity and strong convexity, strong quasar-convexity adds a quadratic growth term that ensures unique minimizers and enables linear convergence rates.

Definition 5.17 (Strongly Quasar-Convex Function)**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ be a convex set and let $f : \mathcal{X} \to \mathbb{R}$ be continuously differentiable on $\mathcal{X}$. Suppose that the set of global minimizers $\mathcal{X}^\star$ is nonempty, and fix $x^\star \in \mathcal{X}^\star$. For parameters $\gamma \in (0, 1]$ and $\mu \geq 0$, the function f is (μ, γ)*-strongly quasar-convex on* $\mathcal{X}$ *at* $x^\star$ if for all $x \in \mathcal{X}$,

$$f(x) - f(x^\star) \leq \frac{1}{\gamma} \nabla f(x)^\top (x - x^\star) - \frac{\mu}{2} \|x - x^\star\|^2.$$

If $\mu > 0$, then $x^\star$ is the unique global minimizer of f on $\mathcal{X}$. □

When $\gamma = 1$ and $\mu > 0$, strong quasar-convexity reduces to strong star-convexity (Definition 5.5).

Property 5.18 (Uniqueness of the Minimizer)**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ be convex and $f : \mathcal{X} \to \mathbb{R}$ be continuously differentiable. If f is (μ, γ)-strongly quasar-convex on $\mathcal{X}$ for some $\gamma \in (0, 1]$ and $\mu > 0$, then f admits a unique global minimizer in $\mathcal{X}$. □

Property 5.19 (Equivalent Characterization of Strong Quasar-Convexity)**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ be a convex set, $f : \mathcal{X} \to \mathbb{R}$ be differentiable, and fix $x^\star \in \mathcal{X}$. For $\gamma \in (0, 1]$ and $\mu \geq 0$, the function f is (μ, γ)-strongly quasar-convex on $\mathcal{X}$ at $x^\star$ if and only if for all

$x \in \mathcal{X}$ and $\lambda \in [0,1]$,

$$f\big(\lambda x^\star + (1-\lambda)x\big) + \lambda\left(1 - \frac{\lambda}{2-\gamma}\right)\frac{\gamma\mu}{2}\,\|x^\star - x\|^2 \;\le\; \gamma\lambda\, f(x^\star) + (1-\gamma\lambda)\, f(x).$$

In this case, $x^\star$ is a global minimizer of f on $\mathcal{X}$; moreover, if $\mu > 0$, then $x^\star$ is the unique global minimizer. □

5.2.4. Projected quasar-convexity

In applications involving constrained optimization and stochastic optimization with nonconvex objectives, a natural variant of quasar-convexity arises where the reference point is not a fixed global minimizer but rather the projection onto the set of all global minimizers. This notion, termed *projected quasar-convexity*, has recently gained attention in the analysis of first-order stochastic methods [84].

Definition 5.20 (Projected Quasar-Convex Function)**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ be a convex set and let $f : \mathcal{X} \to \mathbb{R}$ be continuously differentiable on $\mathcal{X}$. Suppose that the set of global minimizers $\mathcal{X}^\star$ is nonempty, and denote the optimal value by $f^\star := \min_{x\in\mathcal{X}} f(x)$. For $\gamma \in (0,1]$, the function f is *projected γ-quasar-convex on $\mathcal{X}$* if for all $x \in \mathcal{X}$,

$$f(x) - f^\star \le \frac{1}{\gamma}\,\nabla f(x)^\top \big(x - x^p\big),$$

where $x^p := \mathrm{Proj}_{\mathcal{X}^\star}(x)$ denotes the Euclidean projection of x onto $\mathcal{X}^\star$. □

Remark 5 (Comparison with Standard Quasar-Convexity)**.** The key distinction is that projected quasar-convexity requires the inequality to hold with respect to the projection x^p onto the minimizer set $\mathcal{X}^\star$, rather than a fixed minimizer $x^\star \in \mathcal{X}^\star$. When $\mathcal{X}^\star$ is a singleton, the two definitions coincide. However, when $\mathcal{X}^\star$ contains multiple points, projected quasar-convexity is a stronger condition.

Property 5.21 (Projected Quasar-Convexity Implies Standard Quasar-Convexity)**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ be a convex set and $f : \mathcal{X} \to \mathbb{R}$ be continuously differentiable with convex set of global minimizers $\mathcal{X}^\star$. If f is projected γ-quasar-convex on $\mathcal{X}$ for some $\gamma \in (0,1]$, then f is γ-quasar-convex with respect to every point $x^\star \in \mathcal{X}^\star$. □

Definition 5.22 (Projected Strongly Quasar-Convex Function)**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ be a convex set and $f : \mathcal{X} \to \mathbb{R}$ be continuously differentiable. For parameters $\gamma \in (0,1]$ and $\mu \ge 0$, the function f is *projected (μ,γ)-strongly quasar-convex on $\mathcal{X}$* if for all $x \in \mathcal{X}$:

$$f(x) - f^\star \le \frac{1}{\gamma}\nabla f(x)^\top (x - x^p) - \frac{\mu}{2}\|x - x^p\|^2,$$

where $x^p = \mathrm{Proj}_{\mathcal{X}^\star}(x)$. □

Analogous to Property 5.21, projected strong quasar-convexity implies strong quasar-convexity with respect to every point in $\mathcal{X}^\star$ when the minimizer set is convex.

Projected quasar-convexity is particularly natural in stochastic optimization settings where the exact location of minimizers is unknown. The projected variant allows for analysis of convergence to the minimizer set $\mathcal{X}^\star$ rather than a specific point, which is crucial for understanding the behavior of stochastic gradient methods on nonconvex objectives [84]. Moreover, the projected formulation naturally handles problems with multiple global minimizers without requiring a priori knowledge of their locations.

5.2.5. *Tilted convexity*

Tilted convexity provides a two-sided generalization that implies quasar-convexity and captures additional structure useful for algorithm design [63, 64].

Definition 5.23 (Tilted Convex Function)**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ be a convex set and let $f : \mathcal{X} \to \mathbb{R}$ be continuously differentiable. For parameters $\gamma, \gamma_p \in (0, 1]$, the function f is (γ, γ_p)-*tilted convex* on $\mathcal{X}$ if for all $x, y \in \mathcal{X}$,

$$\begin{cases} f(x) + \dfrac{1}{\gamma} \nabla f(x)^\top (y - x) \ \le \ f(y), & \text{if } \nabla f(x)^\top (y - x) \le 0, \\ f(x) + \gamma_p \nabla f(x)^\top (y - x) \ \le \ f(y), & \text{if } \nabla f(x)^\top (y - x) \ge 0. \end{cases}$$

□

The parameter γ controls the behavior when moving in a descent direction (toward lower function values), while γ_p controls the behavior in ascent directions. When $\gamma = \gamma_p = 1$, tilted convexity reduces to standard convexity.

Property 5.24 (Tilted Convexity Implies Quasar-Convexity)**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ be a convex set and $f : \mathcal{X} \to \mathbb{R}$ be continuously differentiable. If f is (γ, γ_p)-tilted convex on $\mathcal{X}$ for some $\gamma, \gamma_p \in (0, 1]$, then for any first-order stationary point $x^\star \in \mathcal{X}$ (i.e., $\nabla f(x^\star)^\top (y - x^\star) \ge 0$ for all $y \in \mathcal{X}$), the point $x^\star$ is a global minimizer and f is γ-quasar-convex at $x^\star$. □

Property 5.25 (Smoothness Lower Bound for the Class)**.** The class of (μ, γ)-strongly quasar-convex functions (with $\gamma \in (0, 1]$ and $\mu > 0$) is not contained in the class of L-smooth functions for any $L < \frac{\gamma\mu}{2-\gamma}$. That is, for every $L < \frac{\gamma\mu}{2-\gamma}$, there exists a (μ, γ)-strongly quasar-convex function whose gradient is not L-Lipschitz. □

5.2.6. *Geometric properties and closure properties*

Property 5.26 (Star-Convexity of the Minimizer Set)**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ be convex and $f : \mathcal{X} \to \mathbb{R}$ be continuously differentiable and γ-quasar-convex on $\mathcal{X}$ for some $\gamma \in (0, 1]$. Then the set of global minimizers $\mathcal{X}^\star$ is star-convex. That is, for any $x^\star \in \mathcal{X}^\star$, the line segment $[x^\star, x] \subseteq \mathcal{X}^\star$ for all $x \in \mathcal{X}^\star$. □

This property ensures that the solution set has favorable geometric structure, which is important for both theoretical analysis and algorithmic convergence.

Quasar-convexity enjoys several closure properties that facilitate its use in composite optimization problems.

Property 5.27 (Nonnegative Weighted Sums Preserve Quasar-Convexity)**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ be a convex set and $x^\star \in \mathcal{X}$. Suppose $f_1, \cdots, f_k : \mathcal{X} \to \mathbb{R}$ are continuously differentiable and γ-quasar-convex on $\mathcal{X}$ with respect to the common minimizer $x^\star$, for some $\gamma \in (0, 1]$. Then, for any non-negative weights $\lambda_1, \cdots, \lambda_k \ge 0$, the function $f = \sum_{i=1}^k \lambda_i f_i$ is also γ-quasar-convex on $\mathcal{X}$ at $x^\star$. □

Property 5.28 (Stability Under Addition for Strong Quasar-Convexity)**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ be convex and $x^\star \in \mathcal{X}$. Suppose $f, g : \mathcal{X} \to \mathbb{R}$ are continuously differentiable and (μ_1, γ_1)- and (μ_2, γ_2)-strongly quasar-convex on $\mathcal{X}$ with respect to the common minimizer $x^\star$, where $\gamma_1, \gamma_2 \in (0, 1]$ and $\mu_1, \mu_2 \geq 0$. Then the sum $h := f + g$ is $(\mu_1 + \mu_2,\, \gamma)$-strongly quasar-convex at $x^\star$, where $\gamma = \min\{\gamma_1, \gamma_2\}$. □

Property 5.29 (Stability Under Finite Summation)**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ be convex and $x^\star \in \mathcal{X}$. For $i = 1, \cdots, k$, let $f_i : \mathcal{X} \to \mathbb{R}$ be continuously differentiable and (μ_i, γ_i)-strongly quasar-convex on $\mathcal{X}$ with respect to the common minimizer $x^\star$, where $\gamma_i \in (0, 1]$ and $\mu_i \geq 0$. Define $h(x) := \sum_{i=1}^k f_i(x)$. Then h is (μ, γ)-strongly quasar-convex at $x^\star$, where $\gamma = \min_{1 \leq i \leq k} \gamma_i$ and $\mu = \sum_{i=1}^k \mu_i$. □

Property 5.30 (Affine-Scaling Invariance)**.** Let $f : \mathbb{R}^n \to \mathbb{R}$ be γ-quasar-convex on $\mathbb{R}^n$ with respect to a minimizer $x^\star$, where $\gamma \in (0, 1]$. For any scalars $a \geq 0$ and $b \neq 0$, define $g(x) := a\, f(bx)$. Then g is γ-quasar-convex on $\mathbb{R}^n$ with respect to the minimizer $x_g^\star := x^\star / b$. □

Property 5.31 (Scaling of Strong Quasar-Convexity Parameters)**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ be convex and $f : \mathcal{X} \to \mathbb{R}$ be differentiable. Suppose f is (μ, γ)-strongly quasar-convex on $\mathcal{X}$ with respect to a minimizer $x^\star \in \mathcal{X}$, where $\gamma \in (0, 1]$ and $\mu > 0$. Then, for any $\theta \in (0, 1]$, the function f is also $(\mu/\theta,\, \theta\gamma)$-strongly quasar-convex at $x^\star$. □

This scaling property reveals a trade-off between the quasar-convexity parameter γ and the strong convexity modulus μ: one can increase the effective strong convexity at the cost of a smaller γ, and vice versa.

5.2.7. Algorithmic implications and applications

Quasar-convexity has found broad application in the analysis of first-order optimization methods for nonconvex problems. The key insight is that the structure provided by quasar-convexity is sufficient to establish convergence rates comparable to those for convex optimization.

Quasar-convexity has been exploited in the convergence analysis of numerous first-order methods. For gradient descent, [8, 21] establish $O(1/N)$ rates and linear convergence under strong quasar-convexity. Accelerated gradient descent with improved rates leveraging momentum has been analyzed in [41, 42]. Extensions to stochastic gradient descent in noisy gradient settings appear in [36, 46], while accelerated stochastic methods combining acceleration with variance reduction are studied in [32].

Beyond gradient methods, proximal algorithms for handling composite objectives with nonsmooth regularizers are analyzed in [28, 41]. Frank–Wolfe and conditional gradient methods for projection-free constrained optimization are developed in [52, 64]. Conjugate gradient methods exploiting curvature information are studied in [56], and adaptive methods including AdaGrad-type algorithms with adaptive step sizes are analyzed in [59, 89].

Notable applications of quasar-convexity include training of linear dynamical systems [39], neural network optimization in overparameterized regimes [90], and various machine learning settings where the optimization landscape exhibits structured nonconvexity. For additional characterizations of quasar-convexity and further theoretical developments, we refer the reader to [42].

6. Regularity Conditions

Beyond generalized convexity, a variety of *regularity conditions* have emerged as powerful tools for analyzing optimization algorithms on nonconvex landscapes. Notable examples include the Polyak–Łojasiewicz (PL) condition [76], error bounds (EB) [61], quadratic growth (QG) [6], essential strong convexity (ESC) [58], the restricted secant inequality (RSI) [93], and weak (or quasi-) strong convexity [66]. As established in [49], these conditions can be viewed as successive relaxations of strong convexity, extending linear convergence guarantees to broad classes of nonconvex functions.

6.1. The Polyak–Łojasiewicz condition

The Polyak–Łojasiewicz (PL) condition has become a cornerstone for establishing linear convergence in nonconvex optimization [2, 3, 18, 49, 76]. Recent work has explored connections between the PL condition and generalized convexity notions. For instance, [41] studies the relationship between strong quasar-convexity and the PL condition, while [90] analyzes convergence properties of smooth quasar-convex functions satisfying PL or QG.

Definition 6.1 (Polyak–Łojasiewicz Condition)**.** Let $f : \mathcal{X} \to \mathbb{R}$ be differentiable and bounded below. The function f satisfies the *Polyak–Łojasiewicz (PL) condition* with parameter $\mu > 0$ if for all $x \in \mathcal{X}$,

$$\frac{1}{2}\,\|\nabla f(x)\|^2 \;\geq\; \mu\big(f(x) - f^\star\big),$$

where $f^\star := \inf_{y\in\mathcal{X}} f(y)$. □

When $\mathcal{X} \subsetneq \mathbb{R}^n$, this property is sometimes referred to as a *local PL condition* [4]. The PL condition is strictly weaker than strong convexity.

Property 6.2 (Strong Convexity Implies PL)**.** Every μ-strongly convex function satisfies the PL condition with parameter μ. □

While the PL condition does not imply convexity, it does imply invexity.

Property 6.3 (PL Implies Invexity)**.** A function satisfying the PL condition is invex with respect to the mapping $\eta(x, y) = -\frac{1}{2\mu}\nabla f(x)$. □

Property 6.4 (Strong Quasar-Convexity Implies PL)**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ be convex and $f : \mathcal{X} \to \mathbb{R}$ be continuously differentiable and (μ, γ)-strongly quasar-convex on $\mathcal{X}$ with respect to a minimizer $x^\star$, where $\gamma \in (0, 1]$ and $\mu > 0$. Then f satisfies the PL condition with parameter $\mu_{\mathrm{PL}} = \gamma^2\mu$:

$$\frac{1}{2}\|\nabla f(x)\|^2 \;\geq\; \gamma^2\mu\big(f(x) - f^\star\big), \quad \forall x \in \mathcal{X}.$$

□

More recently, the PL condition has been interpreted as a special case of the Łojasiewicz gradient inequality [78].

Definition 6.5 (Łojasiewicz Inequality with Exponent θ)**.** Let $f : \mathcal{X} \to \mathbb{R}$ be continuously differentiable. The function f satisfies the *Łojasiewicz inequality* with exponent $\theta \in [0, 1)$ and parameter $\mu > 0$ if, for all $x \in \mathcal{X}$,

$$\frac{1}{2} \|\nabla f(x)\|^2 \;\geq\; \mu\big(f(x) - f^\star\big)^{2\theta},$$

where $f^\star := \inf_{y \in \mathcal{X}} f(y)$. □

Property 6.6. If f satisfies the Łojasiewicz inequality with exponent $\theta = \frac{1}{2}$, then it satisfies the PL condition with the same constant μ. □

Remark 6 (On the Range of Admissible Exponents)**.** If f satisfies Łojasiewicz inequality for some $\theta \in [0, 1)$, then it also satisfies the inequality for any $\tilde{\theta} \in [\theta, 1)$, possibly with a different constant $\tilde{\mu} > 0$. Moreover, if f is L-smooth and non-constant in a neighborhood of some minimizer $x^\star$, then the inequality cannot hold with any exponent $\theta < \frac{1}{2}$ in that neighborhood. □

Finally, for twice continuously differentiable functions, the PL condition is not substantially more general than strong convexity. As shown in [67], bounded minimizer sets together with the PL condition enforce local strong convexity.

Property 6.7. Let $f : \mathcal{X} \to \mathbb{R}$ be twice continuously differentiable and satisfy the PL condition with parameter $\mu > 0$, and assume the set of global minimizers $\mathcal{X}^\star$ is nonempty and bounded. Then:

(1) f admits a unique global minimizer $x^\star$.
(2) there exists $\alpha > f^\star$ such that f is μ-strongly convex on the sublevel set $\mathrm{lev}_\alpha(f)$.

□

The PL condition has been used to establish linear convergence of the ADMM method [92], difference-of-convex algorithms [1, 3], gradient methods [2], and generic classes of descent algorithms [16].

6.2. Error bounds

Error bounds constitute a powerful regularity condition that quantifies how the norm of the gradient controls the distance to the set of minimizers. Unlike strong convexity, the error bound property does not require global curvature and can hold for certain nonconvex or weakly convex functions—making it instrumental in establishing linear convergence for first-order methods beyond the strongly convex regime [29, 49].

Definition 6.8 (Error Bound Condition)**.** Let $f : \mathcal{X} \to \mathbb{R}$ be differentiable, and suppose that the set of global minimizers $\mathcal{X}^\star$ is nonempty. The function f satisfies the *error bound (EB) condition* with parameter $\mu > 0$ if for all $x \in \mathcal{X}$,

$$\|\nabla f(x)\| \geq \mu \, \|x - x^p\|.$$

□

6.3. Quadratic growth

The quadratic growth condition characterizes the curvature of f near the optimal set without requiring convexity.

Definition 6.9 (Quadratic Growth)**.** A function f satisfies *quadratic growth (QG)* with parameter $\mu > 0$ if for all $x \in \mathcal{X}$,

$$f(x) - f^\star \;\geq\; \frac{\mu}{2}\,\|x - x^p\|^2.$$

If f is additionally convex, this property is referred to as *optimal strong convexity* (OSC) [58] or *semi-strong convexity* [66]. □

Property 6.10 (Quadratic Growth of Strongly Quasar-Convex Functions)**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ be convex and $f : \mathcal{X} \to \mathbb{R}$ be continuously differentiable. Suppose f is (μ, γ)-strongly quasar-convex on $\mathcal{X}$ with respect to a minimizer $x^\star \in \mathcal{X}$, where $\gamma \in (0, 1]$ and $\mu > 0$. Then f satisfies the quadratic growth condition:

$$f(x) \;\geq\; f(x^\star) + \frac{\gamma\mu}{2\,(2-\gamma)}\,\|x^\star - x\|^2, \qquad \forall\, x \in \mathcal{X}.$$

□

Unlike the PL and EB conditions, QG alone does not preclude the existence of non-optimal local minima. However, when combined with appropriate descent properties, QG still enables linear convergence to global minimizers.

6.4. Restricted secant inequality

The restricted secant inequality provides a directional strong convexity property along the path to the optimal set.

Definition 6.11 (Restricted Secant Inequality)**.** A function f satisfies the *restricted secant inequality (RSI)* with parameter $\mu > 0$ if for all $x \in \mathcal{X}$,

$$\nabla f(x)^\top (x - x^p) \;\geq\; \mu\,\|x - x^p\|^2,$$

where $x^p = \mathrm{Proj}_{\mathcal{X}^\star}(x)$ and $\mathcal{X}^\star = \mathrm{argmin}_{y\in\mathcal{X}}\, f(y)$.

When f is additionally convex, this condition is termed *restricted strong convexity (RSC)* [93]. □

The RSI condition ensures that the gradient points sufficiently toward the optimal set, which is crucial for establishing linear convergence rates.

6.5. Essential strong convexity and weak strong convexity

Two intermediate conditions between strong convexity and the PL condition are essential strong convexity and weak strong convexity, introduced in [58].

Definition 6.12 (Essential Strong Convexity)**.** A function f satisfies *essential strong convexity (ESC)* with parameter $\mu > 0$ if for all $x, y \in \mathcal{X}$ such that $x^p = y^p$ (i.e., they project to the same optimal point),

$$f(y) \geq f(x) + \nabla f(x)^\top (y - x) + \frac{\mu}{2}\|y - x\|^2.$$

□

Essential strong convexity requires the strong convexity inequality to hold only among points sharing the same nearest optimizer—strictly weaker than global strong convexity when the solution set is non-singleton.

Definition 6.13 (Weak Strong Convexity)**.** A function f satisfies *weak strong convexity (WSC)* with parameter $\mu > 0$ if for all $x \in \mathcal{X}$,

$$f^\star \geq f(x) + \nabla f(x)^\top (x^p - x) + \frac{\mu}{2}\|x^p - x\|^2,$$

where $x^p = \mathrm{Proj}_{\mathcal{X}^\star}(x)$ and $f^\star = \inf_{y \in \mathcal{X}} f(y)$. □

Weak strong convexity requires the strong convexity inequality to hold only in the direction of the nearest optimizer, making it weaker than ESC but still sufficient for linear convergence.

The regularity conditions studied here are related through a well-known hierarchy. For smooth functions, strong convexity implies error bound–type conditions, including the Polyak–Łojasiewicz (PL) inequality, which in turn imply quadratic growth. Under convexity, several of these conditions are equivalent. We refer to [49] for a detailed and complete characterization.

7. Partial Convexity and Pointwise Convexity

7.1. Biconvexity and multiconvexity

Biconvex functions represent an important class of structured nonconvex functions that arise naturally in numerous optimization problems across machine learning, signal processing, and control theory. While jointly nonconvex in their full argument, these functions exhibit convexity when restricted to subsets of variables, a property that enables the development of efficient alternating optimization algorithms [34, 75, 96].

Definition 7.1 (Biconvex Function)**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ and $\mathcal{Y} \subseteq \mathbb{R}^m$ be convex sets. A function $f : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ is *biconvex* if, for every fixed $y \in \mathcal{Y}$, the function $x \mapsto f(x, y)$ is convex on $\mathcal{X}$, and for every fixed $x \in \mathcal{X}$, the function $y \mapsto f(x, y)$ is convex on $\mathcal{Y}$. □

Equivalently, a function $f\colon \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ is biconvex if it is convex in each block of variables when the other block is held fixed. This blockwise convexity naturally extends to related classes: f is called *biconcave* if both partial functions $x \mapsto f(x, y)$ and $y \mapsto f(x, y)$ are concave for all fixed counterparts; *biaffine* if both partial functions are affine; and *bilinear* if both are linear. Notably, bilinear functions constitute a strict subclass of biaffine functions (those vanishing at the origin), and every biaffine function is simultaneously biconvex and biconcave.

Property 7.2. The class of biconvex functions is closed under nonnegative weighted summation. Specifically, if $f_1, \cdots, f_m \colon \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ are biconvex and $w_1, \cdots, w_m \geq 0$, then the function

$$f(x, y) = \sum_{i=1}^{m} w_i f_i(x, y)$$

is biconvex on $\mathcal{X} \times \mathcal{Y}$. □

In particular, nonnegative scaling preserves biconvexity, and the sum of two biconvex functions is biconvex. Additional closure properties are provided in Appendix C.

In practice, solving a biconvex optimization problem—that is, minimizing a biconvex function $f(x, y)$ over a biconvex (often product-form) feasible set $\mathcal{X} \times \mathcal{Y}$—typically relies on heuristic strategies, the most widely used being *alternating convex search* (ACS), also known as block coordinate descent or alternating minimization. In each iteration, ACS fixes one block of variables (e.g., $y^{(k)}$) and solves the convex subproblem $\min_{x \in \mathcal{X}} f(x, y^{(k)})$; it then fixes the updated $x^{(k+1)}$ and solves $\min_{y \in \mathcal{Y}} f(x^{(k+1)}, y)$. Under mild regularity conditions (e.g., compactness of $\mathcal{X}, \mathcal{Y}$ and continuity of f), the sequence of objective values is nonincreasing and converges, and every limit point of the iterates is a *blockwise stationary point* [34]. However, global optimality is not guaranteed due to the nonconvex nature of the joint problem.

Biconvexity has also been extended to more general settings; see [73]. The concept of biconvexity naturally generalizes to functions involving more than two blocks of variables.

Definition 7.3 (Multiconvex Function)**.** Let $\mathcal{X}_1, \ldots, \mathcal{X}_p$ be convex subsets of Euclidean spaces. A function $f : \mathcal{X}_1 \times \cdots \times \mathcal{X}_p \to \mathbb{R}$ is *multiconvex* if it is convex with respect to each block of variables $x^i \in \mathcal{X}_i$ when all other blocks are held fixed.

Formally, for each $i \in \{1, \ldots, p\}$ and for all fixed $(x^1, \ldots, x^{i-1}, x^{i+1}, \ldots, x^p)$, the function

$$x^i \mapsto f(x^1, \ldots, x^{i-1}, x^i, x^{i+1}, \ldots, x^p)$$

is convex on $\mathcal{X}_i$. □

One may also define multiconvex functions by allowing the variables to be partitioned into multiple blocks, such that the objective is convex in each block when the remaining variables are held fixed. Multi-convex problems appear in domains such as machine learning [85]. Biconvexity has emerged as a key structural property in the mathematical optimization community, with applications in robust optimization [51, 95] and quadratic optimization [50].

7.2. Pointwise and midpoint convexity

Classical convexity requires the convexity inequality to hold for all pairs of points. Pointwise convexity weakens this requirement to behavior relative to a single reference point, while midpoint convexity requires the inequality only at $\lambda = \frac{1}{2}$.

Definition 7.4 (Convexity of a Function at a Point)**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ be star-shaped with respect to a point $x^0 \in \mathcal{X}$, and let $f : \mathcal{X} \to \mathbb{R}$. The function f is said to be *convex* at x^0 *(with respect to $\mathcal{X}$)* if for all $x \in \mathcal{X}$ and all $\lambda \in [0, 1]$,

$$f\big(\lambda x^0 + (1-\lambda)x\big) \le \lambda f(x^0) + (1-\lambda) f(x).$$

The function f is said to be *strictly convex at x^0* if the inequality is strict whenever $x \in \mathcal{X}$ with $x \neq x^0$ and $\lambda \in (0, 1)$. □

Definition 7.5 (Pseudo-Convexity at a Point)**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$, let $x^0 \in \mathcal{X}$, and let $f : \mathcal{X} \to \mathbb{R}$ be differentiable at x^0. The function f is said to be *pseudo-convex at x^0 (with respect to $\mathcal{X}$)* if for all $x \in \mathcal{X}$,

$$\nabla f(x^0)^\top (x - x^0) \ge 0 \Longrightarrow f(x) \ge f(x^0).$$

Equivalently, f is pseudo-convex at x^0 if for all $x \in \mathcal{X}$,

$$f(x) < f(x^0) \Longrightarrow \nabla f(x^0)^\top (x - x^0) < 0.$$

□

Definition 7.6 (Quasi-Convexity at a Point)**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ be star-shaped with respect to a point $x^0 \in \mathcal{X}$, and let $f : \mathcal{X} \to \mathbb{R}$. The function f is said to be *quasi-convex at x^0 (with respect to $\mathcal{X}$)* if for all $x \in \mathcal{X}$ and all $\lambda \in [0, 1]$,

$$f(x) \le f(x^0) \Longrightarrow f\big(\lambda x^0 + (1-\lambda)x\big) \le f(x^0).$$

□

Property 7.7 (Epigraph characterization of pointwise convexity)**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ be nonempty, $x^0 \in \mathcal{X}$, and $f : \mathcal{X} \to \mathbb{R}$. Then f is convex at x^0 (with respect to $\mathcal{X}$) if and only if its epigraph is star-shaped at $\big(x^0, f(x^0)\big)$. □

Property 7.8 (First-Order Condition for Pointwise Convexity)**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ be star-shaped with respect to a point $x^0 \in \mathcal{X}$, and let $f : \mathcal{X} \to \mathbb{R}$ be differentiable at x^0. If f is convex at x^0 (with respect to $\mathcal{X}$), then for all $x \in \mathcal{X}$,

$$f(x) - f(x^0) \ge \nabla f(x^0)^\top (x - x^0).$$

If f is strictly convex at x^0, then the inequality is strict for all $x \in \mathcal{X}$ with $x \neq x^0$. □

Property 7.9 (Second-order condition for pointwise convexity)**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ be open and star-shaped at $x^0 \in \mathcal{X}$, and let $f : \mathcal{X} \to \mathbb{R}$ be twice continuously differentiable at x^0. If f is convex at x^0, then the Hessian $\nabla^2 f(x^0)$ is positive semidefinite. □

Property 7.10 (First-Order Condition for Pointwise Quasi-Convexity)**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ be star-shaped with respect to a point $x^0 \in \mathcal{X}$, and let $f : \mathcal{X} \to \mathbb{R}$ be differentiable at x^0. If f is quasi-convex at x^0 (with respect to $\mathcal{X}$), then for all $x \in \mathcal{X}$,

$$f(x) \le f(x^0) \Longrightarrow \nabla f(x^0)^\top (x - x^0) \le 0.$$

Equivalently, if f is quasi-convex at x^0, then for all $x \in \mathcal{X}$,

$$\nabla f(x^0)^\top (x - x^0) > 0 \Longrightarrow f(x) > f(x^0).$$

□

Property 7.11 (Quasi-convexity implies pseudo-convexity under nonzero gradient)**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ be open, and let $f : \mathcal{X} \to \mathbb{R}$ be continuous on $\mathcal{X}$ and differentiable at $x^0 \in \mathcal{X}$. If f is quasi-convex at x^0 (with respect to $\mathcal{X}$) and $\nabla f(x^0) \neq 0$, then f is pseudo-convex at x^0. □

Property 7.12 (Characterization of pseudo-convexity via stationary points)**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ be open and convex, and let $f : \mathcal{X} \to \mathbb{R}$ be differentiable and quasi-convex on $\mathcal{X}$. Then f is pseudo-convex on $\mathcal{X}$ if and only if every stationary point is a global minimizer; that is,

$$\nabla f(x^0) = 0 \Longrightarrow f(x^0) = \inf_{x \in \mathcal{X}} f(x), \qquad \forall\, x^0 \in \mathcal{X}.$$

□

Midpoint convexity, also known as Jensen convexity [44], constitutes a weaker notion than full convexity. It plays a foundational role in the study of functional inequalities and regularity conditions for convex functions.

Definition 7.13 (Midpoint Convex Function)**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ be a convex set and let $f : \mathcal{X} \to \mathbb{R}$. The function f is *midpoint convex* on $\mathcal{X}$ if for all $x, y \in \mathcal{X}$,

$$f\left(\frac{x+y}{2}\right) \leq \frac{f(x) + f(y)}{2}.$$

□

Property 7.14. Every convex function is midpoint convex. □

Property 7.15. If a midpoint convex function f is continuous on $\mathcal{X}$, then f is convex on $\mathcal{X}$. □

8. Structured Function Classes

Beyond classical convexity, a variety of structured function classes have been introduced to model nonlinear phenomena while preserving useful analytical and algorithmic properties. These classes refine or relax standard convexity in different ways: some strengthen convexity through nonlinear transformations (such as logarithmic or exponential mappings), while others decompose nonconvex functions into convex components or dispense with linear structure altogether. The function classes presented in this section—logarithmically convex functions, exponentially convex functions, difference-of-convex (DC) functions, and abstractly convex functions—play a central role in modern optimization, variational analysis, and information theory. They provide flexible modeling tools and underpin many contemporary algorithms for nonconvex optimization.

8.1. Logarithmic convexity

The notion of logarithmic convexity, as introduced and systematically studied by Klinger [54], captures a strengthening of ordinary convexity: rather than requiring the function itself to lie below its chords, one requires that its logarithm does. Intuitively, this means the function grows (or decays) at an accelerating *multiplicative* rate — its curvature is controlled on the logarithmic scale, reflecting exponential-type behavior in the original variable. This property arises naturally in probability, statistics, and information theory, notably in the analysis of moment-generating functions, partition functions, likelihoods, and entropy-related quantities.

Definition 8.1 (Log-Convex Function)**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ be a convex set. A function $f : \mathcal{X} \to \mathbb{R}_{++}$ is called *logarithmically convex* (or *log-convex*) on $\mathcal{X}$ if for all $x, y \in \mathcal{X}$ and all $\lambda \in [0, 1]$,

$$f\big(\lambda x + (1-\lambda)y\big) \;\le\; \big[f(x)\big]^{\lambda}\,\big[f(y)\big]^{1-\lambda}.$$

Equivalently, f is said to be *strictly log-convex* if the inequality is strict for all $\lambda \in (0, 1)$ and all $x \neq y$. □

A function $f : \mathcal{X} \to \mathbb{R}_{++}$ is called *log-concave* if the reverse inequality holds, i.e.,

$$f\big(\lambda x + (1-\lambda)y\big) \;\ge\; \big[f(x)\big]^{\lambda}\,\big[f(y)\big]^{1-\lambda}, \forall\, x, y \in \mathcal{X},\ \lambda \in [0, 1].$$

We now present equivalent characterizations of log-convexity. These formulations highlight structural, analytic, and geometric facets of the concept.

Property 8.2 (Log-Transform Convexity)**.** A function $f : \mathcal{X} \to \mathbb{R}_{++}$ is log-convex on $\mathcal{X}$ if and only if the composition $\log \circ f$ is convex on $\mathcal{X}$. □

Property 8.3 (Exponential Representation)**.** $f : \mathcal{X} \to \mathbb{R}_{++}$ is log-convex on $\mathcal{X}$ if and only if there exists a convex function $h : \mathcal{X} \to \mathbb{R}$ such that $f = e^{\,h}$ on $\mathcal{X}$. □

Property 8.4 (Power Convexity)**.** $f : \mathcal{X} \to \mathbb{R}_{++}$ is log-convex on $\mathcal{X}$ if and only if f^{α} is convex on $\mathcal{X}$ for every $\alpha > 0$. □

Indeed, it suffices to verify convexity for a single exponent $\alpha > 0$ with $\alpha \neq 0$. The case $\alpha = 1$ recovers ordinary convexity, which is necessary but not sufficient for log-convexity.

Property 8.5 (Reciprocal Log-Concavity)**.** $f : \mathcal{X} \to \mathbb{R}_{++}$ is log-convex on $\mathcal{X}$ if and only if $1/f$ is log-concave on $\mathcal{X}$. □

Property 8.6 (Generalized Jensen Inequality)**.** $f : \mathcal{X} \to \mathbb{R}_{++}$ is log-convex on $\mathcal{X}$ if and only if for any finite collection $x^1, \cdots, x^k \in \mathcal{X}$ and weights $\lambda_1, \cdots, \lambda_k \ge 0$ with $\sum_{i=1}^{k} \lambda_i = 1$,

$$f\left(\sum_{i=1}^{k} \lambda_i x^i\right) \;\le\; \prod_{i=1}^{k} \big(f(x^i)\big)^{\lambda_i}.$$

□

Property 8.7 (Gradient Ratio Monotonicity)**.** Assume $f : \mathcal{X} \to \mathbb{R}_{++}$ is differentiable on an open convex set $\mathcal{X}$. Then f is log-convex on $\mathcal{X}$ if and only if the mapping $x \mapsto \nabla \log f(x) = \nabla f(x)/f(x)$ is monotone, i.e., for all $x, y \in \mathcal{X}$,

$$(x - y)^\top \left[\frac{\nabla f(x)}{f(x)} - \frac{\nabla f(y)}{f(y)} \right] \geq 0.$$

□

Property 8.8 (Log-Gradient Inequality)**.** Assume $f : \mathcal{X} \to \mathbb{R}_{++}$ is differentiable on an open convex set $\mathcal{X}$. Then f is log-convex on $\mathcal{X}$ if and only if for all $x, y \in \mathcal{X}$,

$$(x - y)^\top \frac{\nabla f(y)}{f(y)} \leq \log \frac{f(x)}{f(y)} \leq (x - y)^\top \frac{\nabla f(x)}{f(x)}.$$

□

Property 8.9 (Modified Hessian Condition)**.** Assume $f : \mathcal{X} \to \mathbb{R}_{++}$ is twice continuously differentiable on an open convex set $\mathcal{X}$. Then f is log-convex on $\mathcal{X}$ if and only if for every $x \in \mathcal{X}$,

$$\nabla^2 \log f(x) = \frac{1}{f(x)} \nabla^2 f(x) - \frac{1}{f(x)^2} \nabla f(x) \nabla f(x)^\top \succeq 0,$$

or equivalently,

$$f(x) \nabla^2 f(x) - \nabla f(x) \nabla f(x)^\top \succeq 0.$$

□

Property 8.10. Every log-convex function on a convex set $\mathcal{X}$ is convex on $\mathcal{X}$. The converse does not hold in general.

Property 8.11. Every positive concave function on a convex set $\mathcal{X}$ is log-concave on $\mathcal{X}$. The converse does not hold in general.

Property 8.12. For positive functions on a convex set $\mathcal{X}$, log-convexity is a *strictly stronger* condition than convexity, whereas log-concavity is a *strictly weaker* condition than concavity. □

For further details on this concept, we refer the reader to [54].

8.2. Exponential convexity

Exponentially convex functions, studied in [7, 72], strengthen classical convexity by requiring convexity of the exponential transform e^f. This class finds notable applications in mathematical programming, information theory, and entropy optimization [5, 74].

Definition 8.13 (Exponentially Convex Function)**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ be a convex set. A function $f : \mathcal{X} \to \mathbb{R}$ is *exponentially convex* on $\mathcal{X}$ if for all $x, y \in \mathcal{X}$ and all $\lambda \in [0,1]$,

$$e^{f(\lambda x+(1-\lambda)y)} \leq \lambda e^{f(x)} + (1-\lambda)e^{f(y)}.$$

Equivalently, f is exponentially convex if and only if e^f is convex on $\mathcal{X}$. □

Property 8.14 (Logarithmic Characterization)**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ be a convex set. A function $f : \mathcal{X} \to \mathbb{R}$ is exponentially convex if and only if for all $x, y \in \mathcal{X}$ and all $\lambda \in [0,1]$,

$$f(\lambda x + (1-\lambda)y) \leq \log\big(\lambda e^{f(x)} + (1-\lambda)e^{f(y)}\big).$$

□

Property 8.15 (Relationship to r-Convexity)**.** A function $f : \mathcal{X} \to \mathbb{R}$ is exponentially convex if and only if f is 1-convex in the sense of Definition 3.17. Thus, exponential convexity is a special case of r-convexity with $r = 1$. □

Property 8.16 (Exponential Convexity Implies Convexity)**.** Let $f : \mathcal{X} \to \mathbb{R}$ be exponentially convex on a convex set $\mathcal{X} \subseteq \mathbb{R}^n$. Then f is convex on $\mathcal{X}$. The converse does not hold in general. □

8.3. Difference-of-convex functions

Difference-of-convex (DC) programming, pioneered by [40] and systematically developed by [43] [86], and subsequent authors, provides a universal framework for modeling and analyzing nonconvex optimization problems. The class of DC functions is remarkably expressive: it contains all twice continuously differentiable functions on compact convex domains and is closed under most standard algebraic operations.

Definition 8.17 (DC Function)**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ be a convex set. A function $f \colon \mathcal{X} \to \mathbb{R}$ is said to be *difference-of-convex* (or *DC*) if there exist two convex functions $g, h \colon \mathcal{X} \to \mathbb{R}$ such that

$$f(x) = g(x) - h(x), \qquad \forall\, x \in \mathcal{X}.$$

The pair (g, h) is called a *DC decomposition* of f. □

Remark 7. The DC decomposition of a given function is not unique: for any convex ϕ, $f = (g + \phi) - (h + \phi)$ is also a valid DC decomposition.

Property 8.18 (Universality of DC Functions)**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ be a compact convex set. Every function $f \colon \mathcal{X} \to \mathbb{R}$ that is twice continuously differentiable on $\mathcal{X}$ is DC on $\mathcal{X}$. □

Property 8.19 (Algebraic Closure)**.** Let $f_1, f_2 \colon \mathcal{X} \to \mathbb{R}$ be DC functions on a convex set $\mathcal{X} \subseteq \mathbb{R}^n$. Then:

(1) $f_1 + f_2$ and $f_1 - f_2$ are DC;
(2) αf_1 is DC for any $\alpha \in \mathbb{R}$;
(3) If f_1 is bounded below and f_2 is bounded above, then $f_1 \cdot f_2$ is DC;
(4) If f_1 is DC and $\varphi \colon \mathbb{R} \to \mathbb{R}$ is convex and nondecreasing, then $\varphi \circ f_1$ is DC.

□

Additional closure properties, including preservation under pointwise maximum, minimum, and absolute value operations, are provided in Appendix E.

Property 8.20 (Subdifferential Characterization)**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ be open and convex, and let $f = g - h$ be a DC function on $\mathcal{X}$ with g, h convex and locally Lipschitz. Then f is locally Lipschitz and its Clarke subdifferential satisfies

$$\partial_C f(x) \subseteq \partial g(x) - \partial h(x), \qquad \forall\, x \in \mathcal{X},$$

where $\partial g(x)$ and $\partial h(x)$ denote the convex subdifferentials of g and h, respectively. □

Property 8.21 (DC Structure of Other Function Classes)**.** The following inclusions hold:

(1) Every convex function is DC (take $h \equiv 0$);
(2) Every concave function is DC (take $g \equiv 0$);
(3) Every weakly convex (para-convex) function is DC;
(4) Every twice continuously differentiable function on a compact convex domain is DC;
(5) Every polynomial function is DC;
(6) Every rational function with positive denominator on $\mathcal{X}$ is DC.

□

Property 8.22 (DC Decomposition via Quadratic Regularization)**.** Let $f : \mathbb{R}^n \to \mathbb{R}$ be L-smooth (i.e., ∇f is L-Lipschitz). Then f admits the canonical DC decomposition

$$f(x) = \underbrace{\left(f(x) + \frac{L}{2}\|x\|^2 \right)}_{g \text{ convex}} - \underbrace{\frac{L}{2}\|x\|^2}_{h \text{ convex}}.$$

□

Remark 8 (Algorithmic Relevance)**.** The DC structure enables the *DC Algorithm* (DCA), which linearizes the concave part $-h$ at each iteration and solves a convex subproblem:

$$x^{(k+1)} \in \operatorname*{argmin}_{x \in \mathcal{X}} \big\{ g(x) - h(x^{(k)}) - \nabla h(x^{(k)})^\top (x - x^{(k)}) \big\}.$$

Under mild conditions, DCA generates a sequence with monotonically decreasing objective values and converges to a critical point of f. Variants include proximal DCA, inertial DCA, and stochastic DCA. The difference-of-convex algorithm, also known as the convex–concave procedure. We refer to [3] and the references therein for further details.

8.4. Abstract convexity

Abstract convexity, sometimes referred to as *convexity without linearity*, provides a unifying framework that extends many fundamental results of classical convex analysis beyond linear and affine structures. This theory replaces linear functionals with a general family of elementary (or support) functions, enabling optimization and duality theory in significantly broader settings. We refer to [14, 22, 65, 83] for comprehensive treatments and recent applications.

Definition 8.23 (Abstract Convexity)**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ be a nonempty set, and let $\mathcal{H}$ be a family of real-valued functions on $\mathcal{X}$, called the *elementary functions* or *support functions*. A function $f : \mathcal{X} \to \overline{\mathbb{R}}$ is said to be *abstractly convex with respect to* $\mathcal{H}$ (or $\mathcal{H}$*-convex*) if there exists a subset $\mathcal{G} \subseteq \mathcal{H}$ such that

$$f(x) = \sup_{h \in \mathcal{G}} h(x), \qquad \forall x \in \mathcal{X}.$$

Equivalently, f is $\mathcal{H}$-convex if and only if it coincides with its $\mathcal{H}$*-envelope*, defined by

$$f(x) = \sup\{\, h(x) \mid h \in \mathcal{H},\ h(y) \le f(y) \text{ for all } y \in \mathcal{X} \,\}, \qquad \forall x \in \mathcal{X}. \tag{3}$$

□

Property 8.24. When $\mathcal{H}$ is taken to be the family of all affine functions on $\mathbb{R}^n$, that is,

$$\mathcal{H} = \{\, x \mapsto a^\top x + b \mid a \in \mathbb{R}^n,\ b \in \mathbb{R} \,\},$$

the notion of $\mathcal{H}$-convexity reduces to classical convexity, and the $\mathcal{H}$-envelope coincides with the convex envelope (or closed convex hull) of the function. □

9. The Convexity Zoo

To synthesize the diverse notions introduced throughout this survey, we summarize the relationships among classical convexity, generalized convexity, geometric variants, regularity conditions, and special function classes in what we refer to as the *convexity zoo*.

Figures 1 and 2 present two complementary visualizations of this landscape. Figure 1 depicts set-theoretic containments among major function classes, highlighting strict inclusions that hold in general. Figure 2 emphasizes logical implications and structural relationships, including connections that depend on additional assumptions such as smoothness or convexity of the domain.

10. Conclusion

In this survey, we have presented a comprehensive taxonomy of the *Convexity Zoo*, systematically organizing function classes ranging from classical convexity to modern forms of structured nonconvexity. By consolidating definitions, equivalent characterizations, and hierarchical relationships, we have clarified the landscape of function properties that enable efficient optimization.

Our analysis highlights that, while classical convexity remains the gold standard, broader classes—such as quasar-convexity and functions satisfying the Polyak–Łojasiewicz (PL) condition—provide a powerful framework for explaining why many

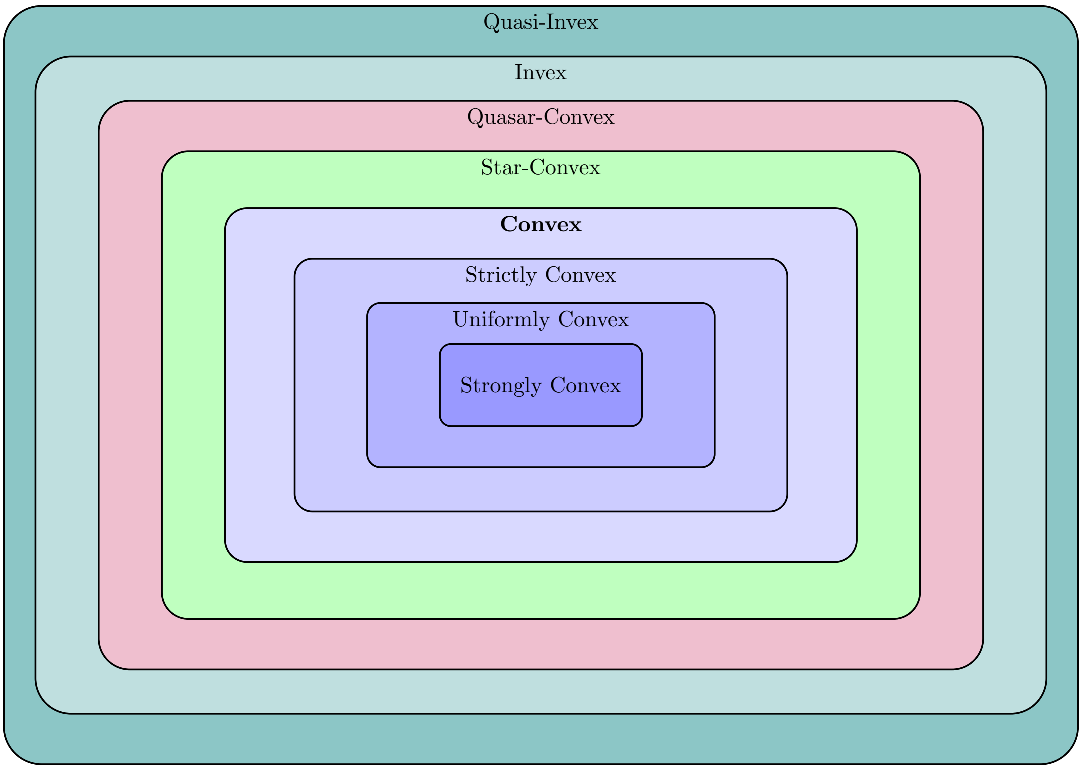

Figure 1. Set-theoretic hierarchy of convexity and generalized convexity classes. Nested regions represent strict inclusions that hold in general, illustrating how classical convexity is embedded within broader nonconvex but structured function classes.

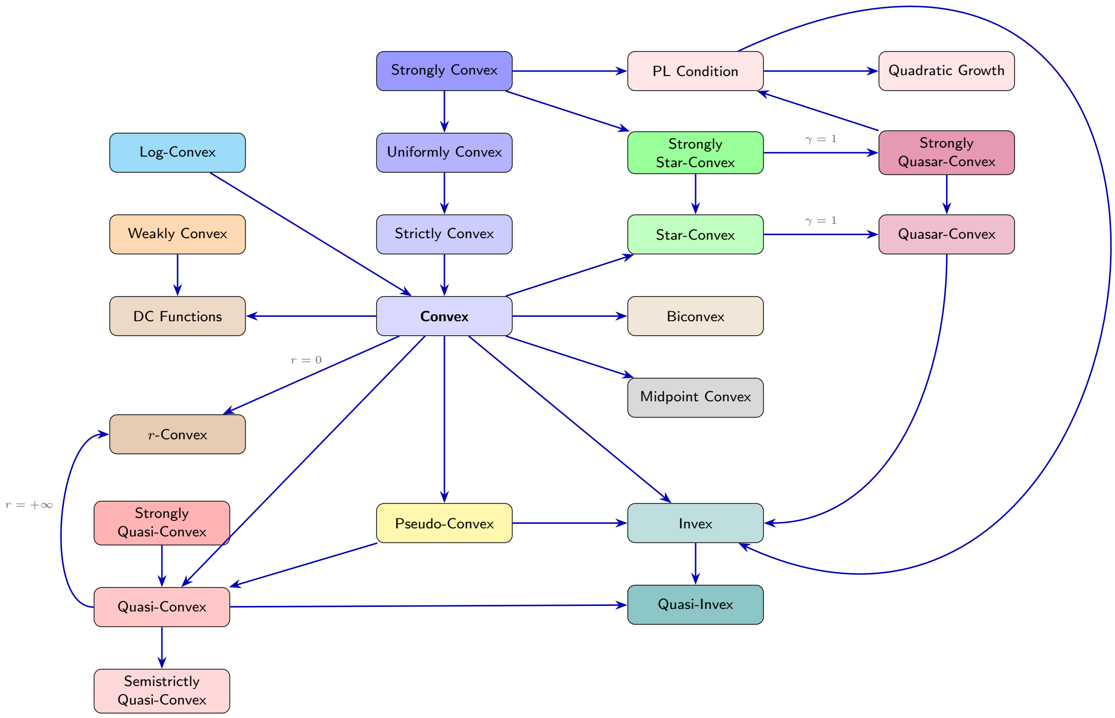

Figure 2. Logical implications and relationships among convexity, generalized convexity, geometric variants, and regularity conditions. Directed arrows indicate implications that hold under standard assumptions (e.g., continuously differentiable and convexity of the domain); absence of an arrow does not imply non-implication.

nonconvex problems arising in machine learning and operations research remain computationally tractable.

We hope that this survey serves as a field guide to the Convexity Zoo, helping researchers navigate, classify, and exploit structure in nonconvex optimization.

Appendix

This appendix complements the main survey by collecting supplementary definitions, equivalent characterizations, and technical properties that support the results in the body of the paper. In particular, it presents additional material on classical convexity (including smoothness and strong convexity variants), refinements and variants of quasi-convexity and generalized convexity, and selected closure/optimality properties referenced in the main text but omitted there for readability.

Appendix A. Supplementary Results on Classical Convexity

This appendix provides additional technical details and properties complementing Section 2.

A.1. Additional Equivalent Characterizations

Property A.1 (Line Restriction)**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ be a convex set and $f : \mathcal{X} \to \mathbb{R}$. Then f is convex if and only if, for every $x, y \in \mathcal{X}$, the univariate function

$$g_{x,y}(t) := f\big((1-t)x + ty\big), t \in [0,1],$$

is convex on the interval $[0, 1]$. □

This characterization reduces multivariate convexity to univariate convexity along all line segments.

Property A.2 (Monotone Secant Slope Criterion)**.** Let $f : \mathcal{X} \to \mathbb{R}$ be a function defined on a convex set $\mathcal{X} \subseteq \mathbb{R}^n$. Then f is convex on $\mathcal{X}$ if and only if, for all distinct points $x, y \in \mathcal{X}$ and for every z strictly between x and y, i.e., $z \in (x, y)$,

$$\frac{f(y) - f(x)}{\|y - x\|} \leq \frac{f(z) - f(y)}{\|z - y\|}.$$

□

Property A.3 (Continuity and Local Lipschitz Regularity)**.** Let $f : \mathcal{X} \to \mathbb{R}$ be a convex function, where $\mathcal{X} \subseteq \mathbb{R}^n$ is an open convex set. Then:

(1) f is continuous on $\mathcal{X}$.
(2) f is locally Lipschitz continuous on $\mathcal{X}$: for every compact subset $K \subseteq \mathcal{X}$, there exists a constant $L_K \geq 0$ such that

$$|f(x) - f(y)| \leq L_K \|x - y\| \text{for all } x, y \in K.$$

□

Property A.4 (Gradient Monotonicity)**.** Let $f : \mathcal{X} \to \mathbb{R}$ be continuously differentiable on a nonempty convex set $\mathcal{X} \subseteq \mathbb{R}^n$. Then f is convex on $\mathcal{X}$ if and only if

$$\big(\nabla f(x) - \nabla f(y)\big)^\top (x - y) \geq 0 \text{for all } x, y \in \mathcal{X}.$$

□

This monotonicity condition is equivalent to saying that ∇f is a *monotone operator*.

A.2. Smoothness

While convexity controls the function from below, smoothness controls it from above by bounding the rate of change of the gradient.

Definition A.5 (L-Smoothness)**.** A differentiable function $f : \mathcal{X} \to \mathbb{R}$ on a convex set $\mathcal{X} \subseteq \mathbb{R}^n$ is *L-smooth* if there exists $L \geq 0$ such that for all $x, y \in \mathcal{X}$

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|.$$

□

Property A.6 (Quadratic Upper Bound)**.** Let $f : \mathcal{X} \to \mathbb{R}$ be convex and L-smooth on an open convex set $\mathcal{X} \subseteq \mathbb{R}^n$. Then for all $x, y \in \mathcal{X}$:

$$f(y) \leq f(x) + \nabla f(x)^\top (y - x) + \frac{L}{2}\|y - x\|^2.$$

□

The quadratic upper bound complements the gradient inequality, providing a "sandwich" for convex smooth functions.

Property A.7 (Equivalent Characterizations of Smoothness)**.** Let $f : \mathcal{X} \to \mathbb{R}$ be continuously differentiable and convex on an open convex set $\mathcal{X} \subseteq \mathbb{R}^n$. The following are equivalent:

(1) f is L-smooth.
(2) (Quadratic upper bound) For all $x, y \in \mathcal{X}$:

$$f(y) \leq f(x) + \nabla f(x)^\top (y - x) + \frac{L}{2}\|y - x\|^2.$$

(3) (Co-coercivity) For all $x, y \in \mathcal{X}$:

$$(\nabla f(x) - \nabla f(y))^\top (x - y) \geq \frac{1}{L}\|\nabla f(x) - \nabla f(y)\|^2.$$

(4) If f is twice differentiable: $\nabla^2 f(x) \preceq LI$ for all $x \in \mathcal{X}$.

□

A.3. Additional Strong Convexity Properties

Property A.8 (Strong Gradient Monotonicity)**.** Let $f : \mathcal{X} \to \mathbb{R}$ be continuously differentiable on a convex set $\mathcal{X} \subseteq \mathbb{R}^n$. Then f is μ-strongly convex if and only if

$$\big(\nabla f(x) - \nabla f(y)\big)^\top (x - y) \geq \mu \, \|x - y\|^2 \text{for all } x, y \in \mathcal{X}.$$

□

Property A.9 (Regularization and Strong Convexity)**.** Let $f : \mathbb{R}^n \to \mathbb{R}$ be a proper function. Then:

(1) f is convex if and only if $f + \frac{\mu}{2}\|\cdot\|^2$ is μ-strongly convex for any $\mu > 0$.
(2) f is μ-strongly convex if and only if $f - \frac{\mu}{2}\|\cdot\|^2$ is convex.

□

A.4. Subdifferential Properties

Property A.10 (Basic Properties of Subdifferentials)**.** Let $f : \mathcal{X} \to \mathbb{R}$ be convex on a convex set $\mathcal{X} \subseteq \mathbb{R}^n$. Then:

(1) $\partial f(x)$ is a closed convex set for each $x \in \mathcal{X}$.
(2) If $x \in \operatorname{int}(\mathcal{X})$, then $\partial f(x)$ is nonempty and bounded.
(3) If f is differentiable at x, then $\partial f(x) = \{\nabla f(x)\}$.
(4) (Sum rule) If $f = f_1 + f_2$ where f_1, f_2 are convex, then $\partial f(x) \supseteq \partial f_1(x) + \partial f_2(x)$, with equality under mild constraint qualifications.

□

A.5. Algebraic Closure Properties

Property A.11 (Nonnegative Weighted Sums)**.** Let $\{f_i\}_{i\in\mathcal{I}}$ be convex functions on a common convex set $\mathcal{X}$, and let $\{\alpha_i\}_{i\in\mathcal{I}}$ be nonnegative scalars. Then $f(x) = \sum_{i\in\mathcal{I}} \alpha_i f_i(x)$ is convex on $\mathcal{X}$. □

Property A.12 (Affine Precomposition)**.** Let $f : \mathbb{R}^m \to \mathbb{R}$ be a convex function, $A \in \mathbb{R}^{m\times n}$, and $b \in \mathbb{R}^m$. Define $g(x) := f(Ax + b)$ on $\mathcal{X} := \{\, x \in \mathbb{R}^n \mid Ax + b \in \operatorname{dom}(f) \,\}$. Then $\mathcal{X}$ is convex and g is convex on $\mathcal{X}$. □

Property A.13 (Composition with Nondecreasing Convex Function)**.** Let $f : \mathcal{X} \to \mathbb{R}$ be convex on a convex set $\mathcal{X} \subseteq \mathbb{R}^n$, and let $g : \mathbb{R} \to \mathbb{R}$ be convex and nondecreasing. Then $h(x) := g\big(f(x)\big)$ is convex on $\mathcal{X}$. □

Property A.14 (General Composition Rules)**.** Let $f : \mathcal{X} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ be twice differentiable. The composition $h = g \circ f$ is convex if either:

(1) g is convex and nondecreasing, and f is convex; or
(2) g is convex and nonincreasing, and f is concave.

Analogously, h is concave if g is concave and nondecreasing with f concave, or g is concave and nonincreasing with f convex. □

Property A.15 (Pointwise Maximum and Supremum)**.** Let $\{f_i\}_{i\in\mathcal{I}}$ be a family of convex functions on a common convex set $\mathcal{X} \subseteq \mathbb{R}^n$. Then the pointwise supremum $f(x) := \sup_{i\in\mathcal{I}} f_i(x)$ is convex on $\mathcal{X}$. □

Property A.16 (Convexity Preserved Under Partial Minimization)**.** Let $f : \mathcal{X}\times\mathcal{Y} \to \mathbb{R}$ be convex, where $\mathcal{X} \subseteq \mathbb{R}^n$ and $\mathcal{Y} \subseteq \mathbb{R}^m$ are convex sets. Define $g(x) := \inf_{y\in\mathcal{Y}} f(x, y)$. Then g is convex on $\mathcal{X}$ (provided $g(x) > -\infty$ for all $x \in \mathcal{X}$). □

Property A.17 (Perspective Transform)**.** Let $f : \mathcal{X} \to \mathbb{R}$ be convex on a convex set $\mathcal{X} \subseteq \mathbb{R}^n$. The *perspective* of f is $g(x, t) := t\, f\left(\frac{x}{t}\right)$ for $t > 0$. Then g is convex on its domain. □

Property A.18 (Closure Properties for Strong Convexity)**.** Let f be μ-strongly convex and g be ν-strongly convex on a common convex set $\mathcal{X}$. Then:

(1) For $\alpha, \beta \geq 0$ with $\alpha + \beta > 0$: $\alpha f + \beta g$ is $(\alpha\mu + \beta\nu)$-strongly convex.
(2) $f + h$ is μ-strongly convex for any convex function h.
(3) If $A \in \mathbb{R}^{n\times m}$ with $\sigma_{\min}(A) > 0$, then $f(Ax)$ is $\mu\sigma_{\min}(A)^2$-strongly convex.

□

A.6. Additional Optimality Conditions

Property A.19 (Convexity of the Solution Set)**.** Let $f : \mathcal{X} \to \mathbb{R}$ be a convex function on a convex set $\mathcal{X} \subseteq \mathbb{R}^n$. Then the set of global minimizers $\operatorname{argmin}_{x\in\mathcal{X}} f(x)$ is convex (and possibly empty). □

Property A.20 (Uniqueness under Strict Convexity)**.** Let $f : \mathcal{X} \to \mathbb{R}$ be strictly convex on a convex set $\mathcal{X} \subseteq \mathbb{R}^n$. If f attains its minimum over $\mathcal{X}$, then the minimizer is unique. □

Remark 9 (Connection to KKT Conditions)**.** For constrained problems $\min_{x\in\mathcal{X}} f(x)$ where $\mathcal{X} = \{x : g_i(x) \leq 0, h_j(x) = 0\}$, the first-order condition extends to the Karush–Kuhn–Tucker (KKT) conditions. Under constraint qualification, $x^\star$ is optimal if and only if there exist multipliers $\lambda_i \geq 0$ and ν_j such that

$$0 \in \partial f(x^\star) + \sum_i \lambda_i \partial g_i(x^\star) + \sum_j \nu_j \nabla h_j(x^\star),$$

along with complementary slackness $\lambda_i g_i(x^\star) = 0$.

Property A.21 (Uniqueness and Quadratic Growth under Strong Convexity)**.** Let $f : \mathcal{X} \to \mathbb{R}$ be μ-strongly convex on a closed convex set $\mathcal{X} \subseteq \mathbb{R}^n$. Then:

(1) If f attains its minimum over $\mathcal{X}$, the minimizer $x^\star$ is unique.
(2) (Quadratic growth) For all $x \in \mathcal{X}$: $f(x) - f(x^\star) \geq \frac{\mu}{2}\|x - x^\star\|^2$.

□

Property A.22 (Two-Sided Bounds for Smooth Strongly Convex Functions)**.** Let $f : \mathcal{X} \to \mathbb{R}$ be μ-strongly convex and L-smooth on a convex set $\mathcal{X} \subseteq \mathbb{R}^n$, with $0 < \mu \leq L$. Let $x^\star$ be the unique global minimizer. Then for all $x \in \mathcal{X}$:

$$\frac{\mu}{2}\|x - x^\star\|^2 \leq f(x) - f(x^\star) \leq \frac{L}{2}\|x - x^\star\|^2.$$

Equivalently, in terms of gradients:

$$\frac{1}{2L}\|\nabla f(x)\|^2 \leq f(x) - f(x^\star) \leq \frac{1}{2\mu}\|\nabla f(x)\|^2.$$

The ratio $\kappa := L/\mu$ is the *condition number* and governs convergence rates of gradient-based methods. □

A.7. *Boundary Behavior and Extreme Points*

Property A.23 (No Interior Maximum)**.** Let $f : \mathcal{X} \to \mathbb{R}$ be a nonconstant convex function on a convex set $\mathcal{X} \subseteq \mathbb{R}^n$. Then f cannot attain a global maximum at any point in the interior of $\mathcal{X}$. □

Property A.24 (Maximum at Extreme Points)**.** Let $f : \mathcal{X} \to \mathbb{R}$ be convex and continuous on a compact convex set $\mathcal{X} \subseteq \mathbb{R}^n$. Then f attains its maximum over $\mathcal{X}$ at some extreme point of $\mathcal{X}$. □

This result underpins the simplex method in linear programming: for linear objectives over polytopes, it suffices to check finitely many vertices.

Appendix B. Supplementary Results on Generalized Convexity

This appendix provides additional variants and technical characterizations complementing Section 3.

B.1. *Additional Quasi-convexity Variants*

Definition B.1 (Strongly Quasi-convex Function)**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ be convex, $f : \mathcal{X} \to \mathbb{R}$, and $\mu > 0$. The function f is *μ-strongly quasi-convex* on $\mathcal{X}$ if for all $x, y \in \mathcal{X}$ and all $\lambda \in [0, 1]$,

$$f\big((1-\lambda)x + \lambda y\big) \leq \max\big\{f(x),\, f(y)\big\} - \frac{\mu}{2}\,\lambda(1-\lambda)\,\|x - y\|^2.$$

□

Definition B.2 (Uniformly Quasi-convex Function)**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ be convex and $f : \mathcal{X} \to \mathbb{R}$. Let $\sigma : \mathbb{R}_+ \to \mathbb{R}_+$ be a function satisfying $\sigma(0) = 0$, $\sigma(t) > 0$ for all $t > 0$, and σ is nondecreasing. Then f is *uniformly quasi-convex on $\mathcal{X}$ with modulus σ* if for all $x, y \in \mathcal{X}$ and all $\lambda \in (0, 1)$,

$$f\big((1-\lambda)x + \lambda y\big) + \lambda(1-\lambda)\,\sigma\big(\|x - y\|\big) \leq \max\big\{f(x),\, f(y)\big\}.$$

□

When $\sigma(t) = \frac{\mu}{2}t^2$, uniform quasi-convexity reduces to μ-strong quasi-convexity.

Definition B.3 (Neatly (or Essentially) Quasi-convex Function)**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ be convex and $f : \mathcal{X} \to \mathbb{R}$. The function f is *neatly quasi-convex* on $\mathcal{X}$ if it is quasi-convex and every local minimizer of f over $\mathcal{X}$ is a global minimizer. □

Definition B.4 (Explicitly Quasi-convex Function)**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ be convex and $f : \mathcal{X} \to \mathbb{R}$. The function f is *explicitly quasi-convex* on $\mathcal{X}$ if it is both quasi-convex and semistrictly quasi-convex on $\mathcal{X}$. □

Definition B.5 (Quasiconcave Function)**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ be convex and $f : \mathcal{X} \to \mathbb{R}$. The function f is *quasiconcave* on $\mathcal{X}$ if $-f$ is quasi-convex on $\mathcal{X}$. Equivalently, for all $x, y \in \mathcal{X}$ and all $\lambda \in [0, 1]$,

$$f\big((1-\lambda)x + \lambda y\big) \geq \min\big\{f(x),\ f(y)\big\}.$$

□

B.2. Additional Characterizations of Quasi-convexity

Property B.6 (Raywise Quasi-convexity)**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ be convex and $f : \mathcal{X} \to \mathbb{R}$. Then f is quasi-convex on $\mathcal{X}$ if and only if for every $x \in \mathcal{X}$ and every direction $d \in \mathbb{R}^n$, the univariate function

$$g_{x,d}(t) := f(x + td), \qquad t \in T_{x,d} := \{\, t \in \mathbb{R} \mid x + td \in \mathcal{X} \,\},$$

is quasi-convex on the (convex) interval $T_{x,d} \subseteq \mathbb{R}$. □

Property B.7 (Segmentwise Quasi-convexity)**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ be convex and $f : \mathcal{X} \to \mathbb{R}$. Then f is quasi-convex on $\mathcal{X}$ if and only if for every $x, y \in \mathcal{X}$, the univariate function

$$h_{x,y}(\lambda) := f\big((1-\lambda)x + \lambda y\big), \qquad \lambda \in [0, 1],$$

is quasi-convex on $[0, 1]$. □

Property B.8 (Second-Order Characterization of Quasi-convexity)**.** Let $\mathcal{X} \subseteq \mathbb{R}^n$ be open and convex, and let f be twice continuously differentiable on $\mathcal{X}$. Suppose that $\nabla f(x) \neq 0$ for all $x \in \mathcal{X}$. Then f is quasi-convex on $\mathcal{X}$ if and only if, for every $x \in \mathcal{X}$, the *bordered Hessian matrix*

$$\widetilde{H}_f(x) := \begin{pmatrix} 0 & \nabla f(x)^\top \\ \nabla f(x) & \nabla^2 f(x) \end{pmatrix} \in \mathbb{R}^{(n+1)\times(n+1)}$$

has the property that its leading principal minors of order $k = 2, 3, \cdots, n+1$ satisfy

$$(-1)^k \det\big(\widetilde{H}_f(x)_{[k]}\big) \ \geq\ 0,$$

where $\widetilde{H}_f(x)_{[k]}$ denotes the $k \times k$ leading principal supmatrix of $\widetilde{H}_f(x)$.

Equivalently, for each $x \in \mathcal{X}$ and for all $d \in \mathbb{R}^n$ such that $\nabla f(x)^\top d = 0$,

$$d^\top \nabla^2 f(x)\, d \ \geq\ 0.$$

□

This second-order characterization states that quasi-convexity requires the Hessian to be positive semidefinite only on the hyperplane orthogonal to the gradient—a weaker condition than positive semidefiniteness everywhere required for convexity.

Appendix C. Supplementary Results on Biconvexity

This appendix provides additional technical details and properties complementing Section 7 on biconvex and multiconvex functions.

Remark 10 (Biconvex sets under weaker assumptions)**.** The notion of a biconvex set is sometimes introduced under weaker structural assumptions than full convexity of the domains $\mathcal{X} \subseteq \mathbb{R}^n$ and $\mathcal{Y} \subseteq \mathbb{R}^m$. In particular, a set $B \subseteq \mathcal{X} \times \mathcal{Y}$ is called *biconvex* if for each fixed $y \in \mathcal{Y}$, the section

$$B_y = \{x \in \mathcal{X} \mid (x, y) \in B\}$$

is convex in $\mathbb{R}^n$, and for each fixed $x \in \mathcal{X}$, the section

$$B_x = \{y \in \mathcal{Y} \mid (x, y) \in B\}$$

is convex in $\mathbb{R}^m$. This formulation — the original one in, e.g., [34] — does not require $\mathcal{X}$ or $\mathcal{Y}$ themselves to be convex; only the vertical and horizontal slices of B must be convex.

Under this definition, biconvexity is strictly weaker than convexity: a biconvex set need not even be connected. Moreover, arbitrary intersections of biconvex sets remain biconvex. □

The pointwise supremum of an arbitrary collection of biconvex functions is biconvex.

Property C.1. The class of biconvex functions is closed under pointwise maximum and supremum. Specifically, if $f_1, \cdots, f_m \colon \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ are biconvex, then the function

$$f(x, y) = \max\{f_1(x, y), \cdots, f_m(x, y)\}$$

is biconvex on $\mathcal{X} \times \mathcal{Y}$. More generally, for any indexed family $\{f_i\}_{i \in I}$ of biconvex functions (where I is an arbitrary index set), the pointwise supremum

$$f(x, y) = \sup_{i \in I} f_i(x, y)$$

is biconvex, provided the supremum is finite for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$. □

Property C.2. Let $\mathcal{X} \subseteq \mathbb{R}^n$ and $\mathcal{Y} \subseteq \mathbb{R}^m$ be nonempty convex sets. Suppose $h \colon \mathbb{R}^p \to \mathbb{R}$ is convex and $g \colon \mathcal{X} \times \mathcal{Y} \to \mathbb{R}^p$ is biaffine. Then the composition

$$f(x, y) := h(g(x, y))$$

is biconvex on $\mathcal{X} \times \mathcal{Y}$. □

Definition C.3. Let $\mathcal{X} \subseteq \mathbb{R}^n$ and $\mathcal{Y} \subseteq \mathbb{R}^m$ be convex sets. A set $\mathcal{B} \subseteq \mathcal{X} \times \mathcal{Y}$ is biconvex if and only if for any $(x^1, y^1), (x^1, y^2), (x^2, y^1), (x^2, y^2) \in \mathcal{B}$ and any $(\lambda, \mu) \in [0, 1] \times [0, 1]$, the point

$$(x^\lambda, y^\mu) := \big((1 - \lambda)x^1 + \lambda x^2,\ (1 - \mu)y^1 + \mu y^2\big)$$

belongs to $\mathcal{B}$. □

Property C.4. Let $\mathcal{X} \subseteq \mathbb{R}^n$ and $\mathcal{Y} \subseteq \mathbb{R}^m$ be nonempty convex sets, and let $f: \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ be biconvex. Then for every $c \in \mathbb{R}$, the sublevel set

$$\mathcal{L}_c := \{(x, y) \in \mathcal{X} \times \mathcal{Y} \mid f(x, y) \leq c\}$$

is a biconvex subset of $\mathcal{X} \times \mathcal{Y}$. □

Property C.5. Let $\mathcal{X} \subseteq \mathbb{R}^n$ and $\mathcal{Y} \subseteq \mathbb{R}^m$ be nonempty convex sets, let $f: \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ be biconvex, and let $\varphi: \mathbb{R} \to \mathbb{R}$ be convex and nondecreasing. Then the composition $h(x, y) = \varphi(f(x, y))$ is biconvex on $\mathcal{X} \times \mathcal{Y}$. □

Appendix D. Supplementary Results on DC Functions

This appendix provides additional technical details and properties complementing Section 8 on difference-of-convex functions.

Property D.1 (Closure Properties of DC Functions)**.** Let $f_1, \ldots, f_n$ be DC functions on a convex set $\mathcal{X}$, and let $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$. Then the following functions are DC:

(1) The linear combination $\lambda_1 f_1(x) + \cdots + \lambda_n f_n(x)$;
(2) The pointwise maximum $\max\{f_1(x), \ldots, f_n(x)\}$;
(3) The pointwise minimum $\min\{f_1(x), \ldots, f_n(x)\}$;
(4) The pointwise product $\prod_{i=1}^n f_i(x)$.

□

Property D.2 (Elementary DC Transformations)**.** Let $f: \mathcal{X} \to \mathbb{R}$ be a DC function on a convex set $\mathcal{X}$. Then the following functions are DC:

(1) The positive part $\max\{0, f(x)\}$;
(2) The negative part $\min\{0, f(x)\}$;
(3) The absolute value $|f(x)|$.

□

Property D.3 (Extended Algebraic Closure)**.** If f_1, f_2 are DC on $\mathcal{X}$, then so are:

(1) $\alpha f_1 + \beta f_2$ for any $\alpha, \beta \in \mathbb{R}$;
(2) $\max\{f_1, f_2\}$ and $\min\{f_1, f_2\}$;
(3) $|f_1|$ (and hence $\max\{0, f_1\}$, $\min\{0, f_1\}$);
(4) $f_1 \cdot f_2$, provided one is bounded below and the other above;
(5) $\varphi \circ f_1$, if $\varphi: \mathbb{R} \to \mathbb{R}$ is convex and nondecreasing.

In particular, all polynomials and rational functions with positive denominators on $\mathcal{X}$ are DC. □

Appendix E. Summary Table of Function Classes

To provide a concise reference summarizing the diverse notions discussed throughout this survey, Table E collects the defining inequalities and core properties of the principal convexity, generalized convexity, and regularity classes.

Table E1. Summary of common convexity, generalized convexity, and regularity classes.

Function Class	**Definition**
Convex	$f(\lambda x + (1-\lambda)y) \le \lambda f(x) + (1-\lambda)f(y),\ \forall x, y \in \mathcal{X},\ \lambda \in [0,1]$.
Strictly Convex	$f(\lambda x + (1-\lambda)y) < \lambda f(x) + (1-\lambda)f(y),\ \forall x \neq y,\ \lambda \in (0,1)$.
μ-Strongly Convex	$f(\lambda x + (1-\lambda)y) \le \lambda f(x) + (1-\lambda)f(y) - \frac{\mu}{2}\lambda(1-\lambda)\|x-y\|^2$.
Uniformly Convex	$f(\lambda x + (1-\lambda)y) \le \lambda f(x) + (1-\lambda)f(y) - \lambda(1-\lambda)\,\varphi(\|x-y\|)$,
Quasi-convex	$f(\lambda x + (1-\lambda)y) \le \max\{f(x), f(y)\},\ \forall x, y \in \mathcal{X},\ \lambda \in [0,1]$.
Strictly Quasi-convex	$f(\lambda x + (1-\lambda)y) < \max\{f(x), f(y)\},\ \forall x \neq y,\ \lambda \in (0,1)$.
Semistrictly Quasi-convex	$f(x) < f(y) \implies f(\lambda x + (1-\lambda)y) < f(y),\ \forall \lambda \in (0,1)$.
Pseudo-convex (C^1)	$f(x) < f(y) \implies \nabla f(y)^\top (x-y) < 0,\ \forall x, y \in \mathcal{X}$.
Strictly Pseudo-convex	$x \neq y,\ f(x) \le f(y) \implies \nabla f(y)^\top (x-y) < 0$.
Invex (C^1)	$\exists\, \eta : \mathcal{X} \times \mathcal{X} \to \mathbb{R}^n$ s.t. $f(x) - f(y) \ge \eta(x,y)^\top \nabla f(y),\ \forall x, y$.
Quasi-invex (C^1)	$\exists\, \eta \not\equiv 0$ s.t. $\eta(x,y)^\top \nabla f(y) > 0 \implies f(x) > f(y),\ \forall x, y$.
Weakly Convex / ρ-Para-convex	$f(\lambda x + (1-\lambda)y) \le \lambda f(x) + (1-\lambda)f(y) + \frac{\rho}{2}\lambda(1-\lambda)\|x-y\|^2$.
(ν, ρ)-Para-convex	$f(\lambda x + (1-\lambda)y) \le \lambda f(x) + (1-\lambda)f(y) + \frac{\rho}{2}\min\{\lambda, 1-\lambda\}\|x-y\|^{1+\nu}$.
Star-Convex (w.r.t. $x^\star$)	$f(\lambda x^\star + (1-\lambda)x) \le \lambda f(x^\star) + (1-\lambda)f(x),\ \forall x,\ \lambda \in [0,1]$.
μ-Strongly Star-Convex (w.r.t. $x^\star$)	$f(\lambda x^\star + (1-\lambda)x) \le \lambda f(x^\star) + (1-\lambda)f(x) - \frac{\mu}{2}\lambda(1-\lambda)\|x-x^\star\|^2$.
γ-Quasar-Convex (C^1, w.r.t. $x^\star$)	$f(x) - f(x^\star) \le \frac{1}{\gamma}\nabla f(x)^\top (x - x^\star),\ \forall x,\ \gamma \in (0,1]$.
(μ, γ)-Strongly Quasar-Convex	$f(x) - f(x^\star) \le \frac{1}{\gamma}\nabla f(x)^\top (x - x^\star) - \frac{\mu}{2}\|x - x^\star\|^2$.
Projected γ-Quasar-Convex	$f(x) - f^\star \le \frac{1}{\gamma}\nabla f(x)^\top (x - x^p), \quad x^p = \mathrm{Proj}_{\mathcal{X}^\star}(x)$.
(γ, γ_p)-Tilted Convex (C^1)	For all $x, y \in \mathcal{X}$, $\begin{cases} f(x) + \frac{1}{\gamma}\nabla f(x)^\top (y-x) \le f(y), & \text{if } \nabla f(x)^\top (y-x) \le 0, \\ f(x) + \gamma_p \nabla f(x)^\top (y-x) \le f(y), & \text{if } \nabla f(x)^\top (y-x) \ge 0. \end{cases}$
PL (Polyak–Łojasiewicz)	$\frac{1}{2}\|\nabla f(x)\|^2 \ge \mu\big(f(x) - f^\star\big),\ \forall x$.
Error Bound (EB)	$\|\nabla f(x)\| \ge \mu\,\|x - x^p\|, \quad x^p = \mathrm{Proj}_{\mathcal{X}^\star}(x)$.
Quadratic Growth (QG)	$f(x) - f^\star \ge \frac{\mu}{2}\|x - x^p\|^2, \quad x^p = \mathrm{Proj}_{\mathcal{X}^\star}(x)$.
Restricted Secant Inequality (RSI)	$\nabla f(x)^\top (x - x^p) \ge \mu\|x - x^p\|^2, \quad x^p = \mathrm{Proj}_{\mathcal{X}^\star}(x)$.
Essential Strong Convexity (ESC)	$x^p = y^p \Longrightarrow f(y) \ge f(x) + \nabla f(x)^\top (y-x) + \frac{\mu}{2}\|y-x\|^2$.
Weak Strong Convexity (WSC)	$f^\star \ge f(x) + \nabla f(x)^\top (x^p - x) + \frac{\mu}{2}\|x^p - x\|^2$.
Biconvex	$f(x,y)$ is convex in x for fixed y and convex in y for fixed x.
Multiconvex	$f(x^1, \ldots, x^p)$ is convex in each block x^i when the others are fixed.
Midpoint Convex	$f\big(\frac{x+y}{2}\big) \le \frac{f(x)+f(y)}{2},\ \forall x, y \in \mathcal{X}$.
Log-Convex	$f > 0$ and $\log f$ is convex on $\mathcal{X}$.
Exponentially Convex	$e^{f(\lambda x + (1-\lambda)y)} \le \lambda e^{f(x)} + (1-\lambda)e^{f(y)},\ \forall x, y,\ \lambda \in [0,1]$.
r-Convex	$f\big(\lambda x + (1-\lambda)y\big) \le \log\big(\lambda(e^{f(x)})^r + (1-\lambda)(e^{f(y)})^r\big)^{1/r}$
Difference-of-Convex (DC)	$\exists\, g, h$ convex s.t. $f = g - h$.
Abstractly Convex ($\mathcal{H}$-convex)	$\exists\, \mathcal{G} \subseteq \mathcal{H}$ s.t. $f(x) = \sup_{h \in \mathcal{G}} h(x),\ \forall x$.

Notes. C^1 denotes continuously differentiable functions. $x^\star$ and x^p denote a global minimizer and the projection onto the minimizer set, respectively. All statements are subject to the assumptions specified in the corresponding sections.

References

[1] H. Abbaszadehpeivasti, E. de Klerk, and A. Taylor, *On the convergence rate of the boosted difference-of-convex algorithm (dca)*, arXiv preprint arXiv:2510.16569 (2025).

[2] H. Abbaszadehpeivasti, E. de Klerk, and M. Zamani, *Conditions for linear convergence of the gradient method for non-convex optimization*, Optimization Letters 17 (2023), pp. 1105–1125.

[3] H. Abbaszadehpeivasti, E. de Klerk, and M. Zamani, *On the rate of convergence of the difference-of-convex algorithm (DCA)*, Journal of Optimization Theory and Applications 202 (2024), pp. 475–496.

[4] A. Aich, A.B. Aich, and B. Wade, *From sublinear to linear: Fast convergence in deep networks via locally polyak-lojasiewicz regions*, arXiv preprint arXiv:2507.21429 (2025).

[5] G. Alirezaei and R. Mathar, *On exponentially concave functions and their impact in information theory*, in *2018 Information Theory and Applications Workshop (ITA)*. IEEE, 2018, pp. 1–10.

[6] M. Anitescu, *Degenerate nonlinear programming with a quadratic growth condition*, SIAM Journal on Optimization 10 (2000), pp. 1116–1135.

[7] T. Antczak, *(p, r)-invex sets and functions*, Journal of mathematical analysis and applications 263 (2001), pp. 355–379.

[8] V. Apidopoulos, J.F. Aujol, C. Dossal, and A. Rondepierre, *Convergence rates of an inertial gradient descent algorithm under growth and flatness conditions*, Mathematical Programming 187 (2021), pp. 151–193.

[9] K.J. Arrow and A.C. Enthoven, *Quasi-concave programming*, Econometrica: Journal of the Econometric Society (1961), pp. 779–800.

[10] F. Atenas, C. Sagastizábal, P.J. Silva, and M. Solodov, *A unified analysis of descent sequences in weakly convex optimization, including convergence rates for bundle methods*, SIAM Journal on Optimization 33 (2023), pp. 89–115.

[11] M. Avriel, *r-convex functions*, Mathematical Programming 2 (1972), pp. 309–323.

[12] A. Barik, S. Sra, and J. Honorio, *Invex programs: First order algorithms and their convergence*, arXiv preprint arXiv:2307.04456 (2023).

[13] H.H. Bauschke, P.L. Combettes, W.M. Menger, and W.M.J. Wouters, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics Vol. 408, Springer, New York, NY, 2011.

[14] E.M. Bednarczuk and M. Syga, *Lagrangian duality for nonconvex optimization problems with abstract convex functions*, arXiv preprint arXiv:2011.09194 (2020).

[15] A. Ben-Israel and B. Mond, *What is invexity?*, The ANZIAM Journal 28 (1986), pp. 1–9.

[16] G. Bento, B. Mordukhovich, T. Mota, and Y. Nesterov, *Convergence of descent optimization algorithms under polyak-łojasiewicz-kurdyka conditions*, Journal of Optimization Theory and Applications 207 (2025), p. 41.

[17] A. Böhm and S.J. Wright, *Variable smoothing for weakly convex composite functions*, Journal of Optimization Theory and Applications 188 (2021), pp. 628–649.

[18] J. Bolte, T.P. Nguyen, J. Peypouquet, and B.W. Suter, *From error bounds to the complexity of first-order descent methods for convex functions*, Mathematical Programming 165 (2017), pp. 471–507.

[19] J.M. Borwein, J.D. Vanderwerff, *et al.*, *Convex functions: constructions, characterizations and counterexamples*, Vol. 109, Cambridge University Press Cambridge, 2010.

[20] S.P. Boyd and L. Vandenberghe, *Convex optimization*, Cambridge University Press, 2004.

[21] J. Bu and M. Mesbahi, *A note on nesterov's accelerated method in nonconvex optimization: a weak estimate sequence approach*, arXiv preprint arXiv:2006.08548 (2020).

[22] H.T. Bui, R.S. Burachik, A.Y. Kruger, and D.T. Yost, *Zero duality gap conditions via abstract convexity*, Optimization 71 (2022), pp. 811–847.

[23] A. Cambini and L. Martein, *Generalized convexity and optimization: Theory and applications*, Springer, 2009.

[24] P. Cannarsa and C. Sinestrari, *Semiconcave Functions, Hamilton—Jacobi Equations, and*

Optimal Control, Springer, 2004.
[25] Y. Chi, Y.M. Lu, and Y. Chen, *Nonconvex optimization meets low-rank matrix factorization: An overview*, IEEE Transactions on Signal Processing 67 (2019), pp. 5239–5269.
[26] A. Choromanska, M. Henaff, M. Mathieu, G.B. Arous, and Y. LeCun, *The loss surfaces of multilayer networks*, in *Artificial Intelligence and Statistics*. PMLR, 2015, pp. 192–204.
[27] D. Davis and B. Grimmer, *Proximally guided stochastic subgradient method for nonsmooth, nonconvex problems*, SIAM Journal on Optimization 29 (2019), pp. 1908–1930.
[28] J. de Brito, F. Lara, and D. Liu, *Extending linear convergence of the proximal point algorithm: The quasar-convex case*, arXiv preprint arXiv:2509.04375 (2025).
[29] D. Drusvyatskiy and A.S. Lewis, *Error bounds, quadratic growth, and linear convergence of proximal methods*, Mathematics of Operations Research 43 (2018), pp. 919–948.
[30] D. Drusvyatskiy and C. Paquette, *Efficiency of minimizing compositions of convex functions and smooth maps*, Mathematical Programming 178 (2019), pp. 503–558.
[31] M. Fazel, R. Ge, S. Kakade, and M. Mesbahi, *Global convergence of policy gradient methods for the linear quadratic regulator*, in *International Conference on Machine Learning*. PMLR, 2018, pp. 1467–1476.
[32] Q. Fu, D. Xu, and A.C. Wilson, *Accelerated stochastic optimization methods under quasar-convexity*, in *International Conference on Machine Learning*. PMLR, 2023, pp. 10431–10460.
[33] G. Giorgi, A. Guerraggio, and J. Thierfelder, *Mathematics of optimization: smooth and nonsmooth case*, Elsevier, 2004.
[34] J. Gorski, F. Pfeuffer, and K. Klamroth, *Biconvex sets and optimization with biconvex functions: a survey and extensions*, Mathematical Methods of Operations Research 66 (2007), pp. 373–407.
[35] A. Goujon, S. Neumayer, and M. Unser, *Learning weakly convex regularizers for convergent image-reconstruction algorithms*, SIAM Journal on Imaging Sciences 17 (2024), pp. 91–115.
[36] R. Gower, O. Sebbouh, and N. Loizou, *Sgd for structured nonconvex functions: Learning rates, minibatching and interpolation*, in *International Conference on Artificial Intelligence and Statistics*. PMLR, 2021, pp. 1315–1323.
[37] S. Guminov, A. Gasnikov, and I. Kuruzov, *Accelerated methods for weakly-quasi-convex optimization problems*, Computational Management Science 20 (2023), p. 36.
[38] M.A. Hanson, *On sufficiency of the kuhn-tucker conditions*, Journal of Mathematical Analysis and Applications 80 (1981), pp. 545–550.
[39] M. Hardt, T. Ma, and B. Recht, *Gradient descent learns linear dynamical systems*, Journal of Machine Learning Research 19 (2018), pp. 1–44.
[40] P. Hartman, *On functions representable as a difference of convex functions*, Pacific Journal of Mathematics 9 (1959), pp. 707–720.
[41] J. Hermant, J.F. Aujol, C. Dossal, and A. Rondepierre, *Study of the behaviour of nesterov accelerated gradient in a non convex setting: the strongly quasar convex case*, arXiv preprint arXiv:2405.19809 (2024).
[42] O. Hinder, A. Sidford, and N. Sohoni, *Near-optimal methods for minimizing star-convex functions and beyond*, in *Conference on learning theory*. PMLR, 2020, pp. 1894–1938.
[43] J.B. Hiriart-Urruty, *Generalized differentiability/duality and optimization for problems dealing with differences of convex functions*, in *Convexity and Duality in Optimization: Proceedings of the Symposium on Convexity and Duality in Optimization Held at the University of Groningen, The Netherlands June 22, 1984*. Springer, 1985, pp. 37–70.
[44] J.L.W.V. Jensen, *Om konvekse funktioner og uligheder imellem middelvaerdier*, Nyt tidsskrift for matematik 16 (1905), pp. 49–68.
[45] J.L.W.V. Jensen, *Sur les fonctions convexes et les inégalités entre les valeurs moyennes*, Acta mathematica 30 (1906), pp. 175–193.
[46] J. Jin, *On the convergence of first order methods for quasar-convex optimization*, arXiv preprint arXiv:2010.04937 (2020).
[47] P. Joulani, A. György, and C. Szepesvári, *A modular analysis of adaptive (non-) convex*

optimization: Optimism, composite objectives, variance reduction, and variational bounds, Theoretical Computer Science 808 (2020), pp. 108–138.
[48] A. Jourani, *Subdifferentiability and subdifferential monotonicity of-paraconvex functions*, Control and Cybernetics 25 (1996).
[49] H. Karimi, J. Nutini, and M. Schmidt, *Linear convergence of gradient and proximal-gradient methods under the polyak-łojasiewicz condition*, in *Joint European conference on machine learning and knowledge discovery in databases*. Springer, 2016, pp. 795–811.
[50] A. Khademi and A. Marandi, *Quadratic optimization through the lens of adjustable robust optimization*, INFORMS Journal on Computing (2025).
[51] A. Khademi, A. Marandi, and M. Soleimani-Damaneh, *A new dual-based cutting plane algorithm for nonlinear adjustable robust optimization*, Journal of Global Optimization 89 (2024), pp. 559–595.
[52] A. Khademi and A. Silveti-Falls, *Adaptive conditional gradient descent*, arXiv preprint arXiv:2510.11440 (2025).
[53] P.D. Khanh, B.S. Mordukhovich, V.T. Phat, and D.B. Tran, *Inexact proximal methods for weakly convex functions*, Journal of Global Optimization 91 (2025), pp. 611–646.
[54] A. Klinger and O. Mangasarian, *Logarithmic convexity and geometric programming*, Journal of Mathematical Analysis and Applications 24 (1968), pp. 388–408.
[55] K. Kume and I. Yamada, *Minimization of nonsmooth weakly convex function over prox-regular set for robust low-rank matrix recovery*, arXiv preprint arXiv:2509.17549 (2025).
[56] I. Kuruzov and F. Stonyakin, *On some versions of subspace optimization methods with inexact gradient information*, arXiv preprint arXiv:2407.01051 (2024).
[57] J.C. Lee and P. Valiant, *Optimizing star-convex functions*, in *2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE, 2016, pp. 603–614.
[58] J. Liu, S. Wright, C. Ré, V. Bittorf, and S. Sridhar, *An asynchronous parallel stochastic coordinate descent algorithm*, in *International Conference on Machine Learning*. PMLR, 2014, pp. 469–477.
[59] Z. Liu, T.D. Nguyen, A. Ene, and H. Nguyen, *On the Convergence of AdaGrad (Norm) on R ˆ d: Beyond Convexity, Non-Asymptotic Rate and Acceleration*, in *International Conference on Learning Representations*. International Conference on Learning Representations, 2023.
[60] P.L. Loh, *Statistical consistency and asymptotic normality for high-dimensional robust m-estimators*, The Annals of Statistics 45 (2017), pp. 866–896.
[61] Z.Q. Luo and P. Tseng, *Error bounds and convergence analysis of feasible descent methods: a general approach*, Annals of Operations Research 46 (1993), pp. 157–178.
[62] O.L. Mangasarian, *Pseudo-convex functions*, SIAM Journal on Control 3 (1965), pp. 281–290.
[63] D. Martínez-Rubio, *Global Riemannian acceleration in hyperbolic and spherical spaces*, in *International conference on algorithmic learning theory*. PMLR, 2022, pp. 768–826.
[64] D. Martínez-Rubio, *Smooth quasar-convex optimization with constraints*, arXiv preprint arXiv:2510.01943 (2025).
[65] R.D. Millán, N. Sukhorukova, and J. Ugon, *Application and issues in abstract convexity*, arXiv preprint arXiv:2202.09959 (2022).
[66] I. Necoara, Y. Nesterov, and F. Glineur, *Linear convergence of first order methods for non-strongly convex optimization*, Mathematical Programming 175 (2019), pp. 69–107.
[67] A.B. Nejma, *Polyak-{\L} ojasiewicz inequality is essentially no more general than strong convexity for c^2 functions*, arXiv preprint arXiv:2512.05285 (2025).
[68] Y. Nesterov, *Lectures on Convex Optimization*, 2nd ed., Springer Optimization and Its Applications Vol. 137, Springer, 2018.
[69] Y. Nesterov and B.T. Polyak, *Cubic regularization of newton method and its global performance*, Mathematical programming 108 (2006), pp. 177–205.
[70] H.T. Nguyen and M. Théra, *Variational analysis of paraconvex multifunctions*, Journal of Optimization Theory and Applications 193 (2022), pp. 180–218.
[71] A. Nishioka, *Revisiting invex functions: Explicit kernel constructions and applications*,

arXiv preprint arXiv:2510.05523 (2025).
[72] M.A. Noor and K.I. Noor, *On exponentially convex functions*, J. Orisa Math. Soc 975 (2019), p. 2323.
[73] M.A. Noor, K.I. Noor, M.T. Rassias, and W.A. Khan, *Biconvex functions and bivariational inequalities*, in *Mathematical Analysis, Differential Equations and Applications*, World Scientific, 2024, pp. 657–687.
[74] S. Pal and T.K.L. Wong, *Exponentially concave functions and a new information geometry*, The Annals of probability 46 (2018), pp. 1070–1113.
[75] P.M. Pardalos and T.M. Rassias, *Geometry and non-convex optimization* (2025).
[76] B.T. Polyak, *Gradient methods for minimizing functionals*, Zhurnal vychislitel'noi matematiki i matematicheskoi fiziki 3 (1963), pp. 643–653.
[77] M. Rahimi, S. Ghaderi, Y. Moreau, and M. Ahookhosh, *Projected subgradient methods for paraconvex optimization: Application to robust low-rank matrix recovery*, arXiv preprint arXiv:2501.00427 (2024).
[78] Q. Rebjock and N. Boumal, *Fast convergence to non-isolated minima: four equivalent conditions for c 2 functions*, Mathematical Programming 213 (2024), p. 151–199.
[79] M. Renaud, A. Leclaire, and N. Papadakis, *On the moreau envelope properties of weakly convex functions*, arXiv preprint arXiv:2509.13960 (2025).
[80] S. Rolewicz, *On γ-paraconvex multifunctions*, Mathematica Japonica 24 (1979), pp. 293–300.
[81] S. Rolewicz, *On alpha (.)-paraconvex and strongly alpha (.)-paraconvex functions*, Control and Cybernetics 29 (2000), pp. 367–377.
[82] S. Rolewicz, *Paraconvex analysis*, Control and Cybernetics 34 (2005), pp. 951–965.
[83] A.M. Rubinov, *Abstract convexity and global optimization*, Vol. 44, Springer Science & Business Media, 2013.
[84] E.M. Saad, W. Lee, and F. Orabona, *New lower bounds for stochastic non-convex optimization through divergence decomposition*, Proceedings of Machine Learning Research 291 (2025), pp. 1–54.
[85] X. Shen, S. Diamond, M. Udell, Y. Gu, and S. Boyd, *Disciplined multi-convex programming*, in *2017 29th Chinese control and decision conference (CCDC)*. IEEE, 2017, pp. 895–900.
[86] H. Tuy, T. Hoang, T. Hoang, V.n. Mathématicien, T. Hoang, and V. Mathematician, *Convex analysis and global optimization*, Springer, 1998.
[87] W. van Ackooij, F. Atenas, and C. Sagastizábal, *Weak convexity and approximate subdifferentials*, Journal of Optimization Theory and Applications 203 (2024), pp. 1686–1709.
[88] J.P. Vial, *Strong and weak convexity of sets and functions*, Mathematics of Operations Research 8 (1983), pp. 231–259.
[89] B. Wang, H. Zhang, Z. Ma, and W. Chen, *Convergence of adagrad for non-convex objectives: Simple proofs and relaxed assumptions*, in *The Thirty Sixth Annual Conference on Learning Theory*. PMLR, 2023, pp. 161–190.
[90] J.K. Wang and A. Wibisono, *Continuized acceleration for quasar convex functions in non-convex optimization*, arXiv preprint arXiv:2302.07851 (2023).
[91] C. Zalinescu, *Convex analysis in general vector spaces*, World scientific, 2002.
[92] M. Zamani, H. Abbaszadehpeivasti, and E. de Klerk, *The exact worst-case convergence rate of the alternating direction method of multipliers*, Mathematical Programming 208 (2024), pp. 243–276.
[93] H. Zhang and W. Yin, *Gradient methods for convex minimization: better rates under weaker conditions*, arXiv preprint arXiv:1303.4645 (2013).
[94] Y. Zhao, S. Wang, and L. Coladas Uria, *Characterizations of r-convex functions*, Journal of optimization theory and applications 145 (2010), pp. 186–195.
[95] J. Zhen, A. Marandi, D. de Moor, D. den Hertog, and L. Vandenberghe, *Disjoint bilinear optimization: A two-stage robust optimization perspective*, INFORMS Journal on Computing 34 (2022), pp. 2410–2427.
[96] H. Zhu and J. Boedecker, *Disciplined biconvex programming*, arXiv preprint

arXiv:2511.01813 (2025).