

A Courcelle-Type Metatheorem for Rank-Bounded Unconstrained Binary Optimization

MARC HARARY*, Columbia University, USA

We present the first uniform XP exact algorithm for unconstrained binary optimization of quadratic, polynomial, fractional, and other objectives under a single parameter, the differentially affine (DA) rank r . An objective $f : \{0, 1\}^n \rightarrow \mathbb{R}$ has DA rank r if there is a feature map $\psi : \{0, 1\}^n \rightarrow \mathbb{R}^r$ such that each coordinate flip has finite gain $\Delta_{\pm e_i} f(x) = \langle v_{\pm e_i}, \psi(x) \rangle + \beta_{\pm e_i}$. Our algorithm enumerates the $O((2n)^r)$ chambers of the induced hyperplane arrangement and applies a two-sided local-optimality test: a solution exists on a chamber and is unique iff $\text{sign } \Delta_{+e_i} = -\text{sign } \Delta_{-e_i}$ for all i , in which case $x_i^* = 1$ iff $\Delta_{+e_i} > 0$. This yields $n^{O(r)}$ time with $O(n)$ decoding per chamber. The framework uniformly covers a wide range of nonlinear functions, including all rank- r quadratics, low-Waring-rank pseudo-Boolean polynomials, finite products/ratios on positive domains, finite-basis separable sums via explicit lifts, Taylor-series approximations of analytic functions, and compositions of all the foregoing. Applications include Ising spin models, optimal experimental design, portfolio optimization, and robust statistics. Prior to our work, only specialized subcases involving sparsity, convexity, submodularity, etc. were known to be tractable. Analogous in spirit to Courcelle’s theorem (MSO on bounded treewidth graphs) and Grohe’s meta-theorems for constraint satisfaction, our result replaces logical width with analytic rank for nonlinear pseudo-Boolean optimization.

ACM Reference Format:

Marc Harary. 2025. A Courcelle-Type Metatheorem for Rank-Bounded Unconstrained Binary Optimization. 1, 1 (November 2025), 14 pages. <https://doi.org/10.1145/nnnnnnnn.nnnnnnnn>

1 Introduction

1.1 Problem setting

Many important NP-hard optimization problems consist of a nonlinear objective over the Hamming hypercube $\{0, 1\}^n$. Examples include quadratic unconstrained binary optimization (QUBO) [46], polynomial unconstrained binary optimization (PUBO), fractional unconstrained binary optimization (FUBO), nonlinear assignment and matching problems, portfolio optimization, and optimal experimental design.

In many such problems, the gain from flipping a coordinate—the change $f(x \pm e_i) - f(x)$ —depends only on a low-dimensional linear function of the current state x . This occurs in low-rank quadratic and polynomial forms, in fractional objectives after cross-multiplication, and other objectives.

We show that the fundamental structural parameter is not the polynomial degree of f , but its *differential rank*: the dimension of the feature space governing all such directional gains. We identify a single predicate condition under which local optimality corresponds to membership in a low-dimensional hyperplane arrangement, making the discrete

*Preferred email: hararymarc2@gmail.com.

Author’s Contact Information: Marc Harary, Columbia University, Department of Computer Science, New York, NY, USA, mh4654@columbia.edu.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2025 Copyright held by the owner/author(s). Publication rights licensed to ACM.

Manuscript submitted to ACM

search space enumerable in $n^{O(r)}$ time and yielding the first uniform XP tractability result for nonlinear pseudo-Boolean optimization.

1.2 Our contributions

- **Differential affineness (DA rank).** We parameterize unconstrained pseudo-Boolean objectives by DA rank r : there exists $\psi: \{0, 1\}^n \rightarrow \mathbb{R}^r$ with flip gains $\Delta_{\pm e_i} f(x) = \langle v_{\pm e_i}, \psi(x) \rangle + \beta_{\pm e_i}$. This single parameter covers rank- r quadratics, low-Waring polynomials, products/ratios on positive domains, affine/polynomial compositions, finite-basis separable sums, and bilinear/multilinear forms, with explicit lifts for each class.
- **Two-sided local optimality criterion.** We prove a chamberwise optimality test: on any chamber, a locally optimal solution exists iff $\text{sign } \Delta_{+e_i} = -\text{sign } \Delta_{-e_i}$ for all i , with unique decode $x_i^* = \mathbf{1}\{\Delta_{+e_i} > 0\}$. This correctly excludes chambers where any coordinate has gains of the same sign in both directions, resolving the boundary-behavior subtlety.
- **XP enumeration algorithm.** With $m = 2n$ hyperplanes, the arrangement in $\mathbb{R}^r$ has $O((2n)^r)$ chambers. Enumerating them and applying the $O(n)$ optimality test yields an $n^{O(r)}$ exact algorithm with no external oracle calls.
- **Explicit constructions.** We provide closed-form feature maps ψ , covectors $\{v_{\pm e_i}\}$, and offsets $\{\beta_{\pm e_i}\}$ with rank bounds for all major function classes: polynomials (via Waring decomposition), products and ratios, compositions, separable sums, and CP-rank tensor models.
- **Applications.** The framework yields $n^{O(r)}$ exact algorithms for rank- r QUBO, PUBO, and FUBO, as well as DA-rank formulations of problems in trimmed/robust statistics, optimal experimental design, and portfolio optimization.
- **Conceptual significance.** DA rank plays a Courcelle-style width role for nonlinear pseudo-Boolean optimization, reducing global optimization to sign-class enumeration in $\mathbb{R}^r$. This yields the first unified parameterized complexity theory for this domain, analogous to treewidth-based results for constraint satisfaction.

1.3 Roadmap

Section 2 fixes notation and recalls affine and central hyperplane arrangements. Section 3 defines differential affineness and proves the main metatheorem, yielding the $n^{O(r)}$ exact algorithm. Section 4 provides explicit lifts and rank bounds for quadratics, low-Waring-rank polynomials, products/ratios on positive domains, affine/polynomial compositions, finite-basis separable sums, and bilinear/multilinear forms. Section 5 instantiates the framework for trimmed/robust statistics, optimal experimental design (D/A/E-optimality), and portfolio optimization. Section 6 situates DA rank alongside existing parametric programming techniques and Courcelle/Grohe meta-theorems. Section 7 summarizes the consequences of bounded DA rank and outlines extensions beyond the unconstrained setting.

2 Preliminaries

2.1 Notation

We write $[n] := \{1, \dots, n\}$ for the index set. For $x \in \{0, 1\}^n$, define its *support* $\text{supp}(x) := \{i : x_i = 1\}$. A *pseudo-Boolean function* is any $f: \{0, 1\}^n \rightarrow \mathbb{R}$. For a feasible perturbation d , the *gain* is

$$\Delta_d f(x) := f(x + d) - f(x).$$

A *feature map* is any function $\psi : \mathbb{R}^n \rightarrow \mathbb{R}^r$.

2.2 Hyperplane arrangements

A *central hyperplane arrangement* in $\mathbb{R}^r$ is a finite set $\mathcal{A} = \{H_i\}_{i=1}^m$ with

$$H_i := \{\xi \in \mathbb{R}^r : w_i^\top \xi = 0\},$$

where $w_i \in \mathbb{R}^r \setminus \{0\}$. A central arrangement defines an *oriented matroid* $\mathcal{M}(\mathcal{A})$ on ground set $[m]$. Each point $\xi \in \mathbb{R}^r$ induces a sign vector $\sigma(\xi) = (\text{sign}(w_1^\top \xi), \dots, \text{sign}(w_m^\top \xi)) \in \{-, 0, +\}^m$, and the set of all such sign vectors forms the covector set of $\mathcal{M}(\mathcal{A})$. The poset of covectors ordered by sign-refinement is isomorphic to the *face lattice* $\mathcal{F}(\mathcal{A})$. The following classical result controls the combinatorial complexity of arrangements and zonotopes:

THEOREM 1 (ZASLAVSKY [52]). *Let $\mathcal{A}$ be a central arrangement of m hyperplanes in $\mathbb{R}^r$. Then*

$$|\mathcal{F}(\mathcal{A})| \leq \sum_{j=0}^r \binom{m}{j} = O(m^r).$$

In particular, the number of maximal top-dimensional faces (known as chambers or topes) is given by $|\mathcal{T}(\mathcal{A})| = O(m^r)$.

3 Differentially affine objectives and chamber-wise readout (unconstrained)

We work on the unconstrained domain $P = \{0, 1\}^n$ with elementary feasible directions (EFDs)

$$\mathcal{D}_1 := \{\pm e_i : i \in [n]\}.$$

For $f : P \rightarrow \mathbb{R}$ and feasible (x, d) with $d \in \mathcal{D}_1$, write the reduced cost

$$\Delta_d f(x) := f(x + d) - f(x).$$

DEFINITION 1 (DIFFERENTIALLY AFFINE (DA) OF RANK r). *We say f is differentially affine of rank r with respect to $\mathcal{D}_1$ (denoted $f \in \mathcal{DA}^r(\mathcal{D}_1)$) if there exist a feature map $\psi : P \rightarrow \mathbb{R}^r$, covectors $v_d \in (\mathbb{R}^r)^*$, and scalars $\beta_d \in \mathbb{R}$ such that*

$$\Delta_d f(x) = \langle v_d, \psi(x) \rangle + \beta_d, \quad \forall x \in P, \forall d \in \mathcal{D}_1 \text{ feasible at } x. \quad (1)$$

The minimal such r is the differential rank $\text{rank}_{\text{DA}}(f; \mathcal{D}_1)$.

Let $\mathcal{A}(\mathcal{D}_1)$ be the affine hyperplane arrangement in $\mathbb{R}^r$ with one hyperplane $H_d := \{\xi \in \mathbb{R}^r : \langle v_d, \xi \rangle + \beta_d = 0\}$ per direction $d \in \mathcal{D}_1$. Write $\mathcal{T}(\mathcal{A})$ for its full-dimensional cells (chambers). For $\xi \notin \bigcup_d H_d$, define the *EFD sign vector*

$$\sigma(\xi) := (\text{sign}(\langle v_d, \xi \rangle + \beta_d))_{d \in \mathcal{D}_1} \in \{-, +\}^{2n},$$

which is constant on each chamber C ; write $\sigma(C)$.

LEMMA 1 (CHAMBER-CONSTANCY OF EFD SIGNS). *If $f \in \mathcal{DA}^r(\mathcal{D}_1)$ and $\psi(x) \in C \in \mathcal{T}(\mathcal{A})$, then for every feasible $d \in \mathcal{D}_1$ the sign $\text{sign } \Delta_d f(x)$ depends only on C and d and equals the d -th entry of $\sigma(C)$.*

PROOF. Immediate from (1). □

3.1 Two-sided toggles, local stability, and exact readout

For $i \in [n]$ and chamber C , introduce the two toggle signs (well-defined by Lemma 1)

$$s_i^+(C) := \text{sign}(\Delta_{+e_i} f(x)), \quad s_i^-(C) := \text{sign}(\Delta_{-e_i} f(x)) \quad (\psi(x) \in C).$$

DEFINITION 2 (TOGGLE-STABILITY IN A CHAMBER). A vector $x^\star \in \{0, 1\}^n$ with $\psi(x^\star) \in C$ is toggle-stable in C if, for every i ,

$$x_i^\star = 0 \Rightarrow \Delta_{+e_i}f(x^\star) \leq 0, \quad x_i^\star = 1 \Rightarrow \Delta_{-e_i}f(x^\star) \leq 0.$$

LEMMA 2 (PER-COORDINATE FEASIBILITY, FORCING, AND AMBIGUITY). Fix a chamber C and assume nondegeneracy $s_i^\pm(C) \in \{\pm 1\}$.

- (1) Feasibility filter: A toggle-stable point with $\psi(x) \in C$ exists if and only if no coordinate has $(s_i^+, s_i^-) = (+, +)$.
- (2) Forced choices: If $(s_i^+, s_i^-) = (+, -)$ then any toggle-stable x must have $x_i = 1$; if $(s_i^+, s_i^-) = (-, +)$ then any toggle-stable x must have $x_i = 0$.
- (3) Ambiguity: If $(s_i^+, s_i^-) = (-, -)$, then both $x_i = 0$ and $x_i = 1$ satisfy Definition 2 at coordinate i .

PROOF. (1) If $(+, +)$ occurs at i , then $\Delta_{\pm e_i}f(x) > 0$ for any x with $\psi(x) \in C$, so no x is stable. Conversely, if each coordinate avoids $(+, +)$, assign x_i by (2) and choose arbitrarily when $(-, -)$; the resulting x meets Definition 2. (2) If $(+, -)$ and $x_i = 0$, then $\Delta_{+e_i}f(x) > 0$ contradicts stability; hence $x_i = 1$ is forced. The case $(-, +)$ is symmetric. (3) If $(-, -)$, then both toggles decrease f , so either value of x_i is locally admissible. $\square$

PROPOSITION 1 (TWO-SIDED READOUT; UNIQUENESS IFF OPPOSITE SIGNS). Fix a chamber C with nondegenerate signs. Then:

- (1) Existence: A toggle-stable point with $\psi(x) \in C$ exists iff the feasibility filter of Lemma 2(1) holds.
- (2) Uniqueness: A toggle-stable point is unique iff $s_i^+(C) = -s_i^-(C)$ for every i . In that case,

$$x_i^\star(C) = \mathbf{1}\{s_i^+(C) = +\} \quad (i \in [n]).$$

- (3) Optimality within C : Any global maximizer x^{opt} with $\psi(x^{\text{opt}}) \in C$ is toggle-stable, since an improving toggle would contradict global optimality. If uniqueness holds, then $x^{\text{opt}} = x^\star(C)$.

PROOF. (1) Follows immediately from Lemma 2(1): if any coordinate has $(s_i^+, s_i^-) = (+, +)$ then no point is toggle-stable; if none do, assigning coordinates per Lemma 2(2) and choosing arbitrarily when $(-, -)$ yields a toggle-stable point.

(2) If some $(s_i^+, s_i^-) = (-, -)$, then both $\Delta_{+e_i}f(x) < 0$ and $\Delta_{-e_i}f(x) < 0$ on C , so both $x_i = 0$ and $x_i = 1$ satisfy Definition 2 at coordinate i , giving at least two distinct toggle-stable points. Conversely, if $s_i^+(C) = -s_i^-(C)$ for all i , then Lemma 2(2) forces $x_i = 1$ when $(s_i^+, s_i^-) = (+, -)$ and $x_i = 0$ when $(-, +)$, yielding the unique decode $x_i^\star(C) = \mathbf{1}\{s_i^+(C) = +\}$.

(3) Let x^{opt} be a global maximizer with $\psi(x^{\text{opt}}) \in C$. If some coordinate admits an improving toggle, then f increases, contradicting global optimality. Hence x^{opt} is toggle-stable. When uniqueness from (2) holds, necessarily $x^{\text{opt}} = x^\star(C)$. $\square$

REMARK (DEGENERACY AND GENERIC PERTURBATION). If some $s_i^\pm(C) = 0$, apply an infinitesimal perturbation $\beta_{\pm e_i} \leftarrow \beta_{\pm e_i} + \varepsilon \rho_{\pm e_i}$ with fixed ρ and $\varepsilon \downarrow 0$ to obtain strict signs without changing the DA rank; the limit selects a maximizer under a lexicographic refinement of f .

3.2 Enumeration complexity and algorithmic guarantee

Here $|\mathcal{D}_1| = 2n$. By Theorem 1, the number of chambers satisfies

$$|\mathcal{T}(\mathcal{A})| = O((2n)^r) = n^{O(r)}.$$

THEOREM 2 (XP META-THEOREM IN DIFFERENTIAL RANK (UNCONSTRAINED)). *Let $f \in \mathcal{DA}^r(\mathcal{D}_1)$. There is an algorithm that enumerates all $O((2n)^r)$ chambers $C \in \mathcal{T}(\mathcal{A})$ and, for each C :*

- (1) *reads the two-sided signs $\{s_i^\pm(C)\}_{i=1}^n$ in $O(n)$ time;*
- (2) *discards C if any $(s_i^+, s_i^-) = (+, +)$;*
- (3) *otherwise outputs the unique toggle-stable $x^*(C)$ when all pairs are opposite, or applies the perturbation scheme of Section 3.1 to break ties.*

Among all chambers that pass the filter, return the best $x^(C)$. The total running time is $O(n \cdot (2n)^r) = n^{O(r)}$.*

4 Objective instantiations

4.1 Quadratic objectives are differentially affine

Let $Q \in \mathbb{R}^{n \times n}$ be symmetric with $\text{rank}(Q) = r$. Fix a rank factorization

$$Q = B^\top B, \quad B \in \mathbb{R}^{r \times n} \text{ full row rank.}$$

Consider $f(x) = x^\top Qx$ on any feasible set $P \subseteq \{0, 1\}^n$ and any finite EFD set $\mathcal{D} \subset \mathbb{Z}^n$ (e.g., $\mathcal{D}_1 = \{\pm e_i\}$ for unconstrained toggles and $\mathcal{D}_2 = \{\pm(e_i - e_j)\}$ for Top- K exchanges).

PROPOSITION 2 (QUADRATICS ARE DA OF RANK r). *Define the feature map $\psi : P \rightarrow \mathbb{R}^r$ by $\psi(x) := Bx$. For each $d \in \mathcal{D}$, set*

$$v_d := 2Bd \in \mathbb{R}^r, \quad \beta_d := d^\top Qd \in \mathbb{R}.$$

Then, for every feasible (x, d) ,

$$\Delta_d f(x) = f(x + d) - f(x) = 2d^\top Qx + d^\top Qd = \langle v_d, \psi(x) \rangle + \beta_d.$$

Hence $f \in \mathcal{DA}^r(\mathcal{D})$ with differential rank equal to $\text{rank}(Q)$.

PROOF. Using $Q = B^\top B$,

$$\Delta_d f(x) = (x + d)^\top Q(x + d) - x^\top Qx = 2d^\top Qx + d^\top Qd = 2(Bd)^\top (Bx) + d^\top Qd = \langle 2Bd, Bx \rangle + \beta_d.$$

This matches the DA form with $\psi(x) = Bx$, $v_d = 2Bd$, and $\beta_d = d^\top Qd$. □

COROLLARY 1 (AFFINE-QUADRATICS REMAIN DA OF RANK r). *For $f(x) = x^\top Qx + c^\top x + \text{const}$, the same ψ and v_d as in Proposition 2 yield*

$$\Delta_d f(x) = \langle v_d, \psi(x) \rangle + \underbrace{(d^\top Qd + c^\top d)}_{\beta_d},$$

so $f \in \mathcal{DA}^r(\mathcal{D})$. The linear term contributes only to the intercepts $\{\beta_d\}$ and does not increase differential rank.

This subsumes Ising [27] and Potts [45] models with low-rank coupling; pairwise Markov random field [50] and conditional random field [31] energies; and Hopfield networks [31] with low-rank Q .

4.2 Polynomial objectives

Let $f : \{0, 1\}^n \rightarrow \mathbb{R}$ be a pseudo-Boolean polynomial of degree at most D ,

$$f(x) = \sum_{t=0}^D \langle \mathcal{T}^{(t)}, x^{\otimes t} \rangle,$$

where each $\mathcal{T}^{(t)}$ is a symmetric order- t tensor. Assume a symmetric CP/Waring factorization [29]

$$\mathcal{T}^{(t)} = \sum_{\ell=1}^{k_t} \alpha_{t\ell} \left(u^{(t,\ell)} \right)^{\otimes t}, \quad \alpha_{t\ell} \in \mathbb{R}, u^{(t,\ell)} \in \mathbb{R}^n.$$

Define the base linear forms $s_{t\ell}(x) := \langle u^{(t,\ell)}, x \rangle$ and let

$$r := \sum_{t=1}^D tk_t, \quad \psi(x) := (s_{t\ell}(x)^q)_{\substack{t=1,\dots,D \\ \ell=1,\dots,k_t \\ q=0,\dots,t-1}} \in \mathbb{R}^r.$$

PROPOSITION 3 (LOW-RANK POLYNOMIALS ARE DA). *For any finite EFD set $\mathcal{D}$ and any feasible (x, d) (so $x, x+d \in P$), there exists $w_d \in (\mathbb{R}^r)^*$, independent of x , such that*

$$\Delta_d f(x) = \langle w_d, \psi(x) \rangle.$$

Hence $f \in \mathcal{DA}^r$ with $r = \sum_{t=1}^D tk_t$.

PROOF. Write $f = \sum_{t,\ell} \alpha_{t\ell} s_{t\ell}(x)^t$. For fixed (t, ℓ) ,

$$s_{t\ell}(x+d)^t - s_{t\ell}(x)^t = \sum_{j=1}^t \binom{t}{j} s_{t\ell}(x)^{t-j} \langle u^{(t,\ell)}, d \rangle^j.$$

Summing over (t, ℓ) with coefficients $\alpha_{t\ell}$ yields a linear combination of the coordinates $s_{t\ell}(x)^q$, $q = 0, \dots, t-1$, with d -dependent coefficients; collect them into w_d . $\square$

This subsumes any degree- D pseudo-Boolean with bounded Waring ranks $\{k_t\}$ (e.g., sparse polynomial expansions [3], low-rank higher-order MRFs [50]).

4.3 Closure under products and ratios

Let $\{f_s\}_{s=1}^S$ with $f_s \in \mathcal{DA}^{r_s}$ admit feature maps ψ_s so that for every EFD d ,

$$\Delta_d f_s(x) = \langle w_{s,d}, \psi_s(x) \rangle + \beta_{s,d}.$$

PROPOSITION 4 (FINITE PRODUCTS PRESERVE DA). *Define the lifted feature map*

$$\Psi_\Pi(x) := \bigotimes_{s=1}^S [\psi_s(x); f_s(x); 1] \in \mathbb{R}^{R_\Pi}, \quad R_\Pi \leq \prod_{s=1}^S (r_s + 2).$$

Then for every EFD d there exists $W_d^{(\Pi)}$ with

$$\Delta_d \left(\prod_{s=1}^S f_s \right)(x) = \langle W_d^{(\Pi)}, \Psi_\Pi(x) \rangle,$$

hence $\prod_s f_s \in \mathcal{DA}^{R_\Pi}$.

PROOF. Let $A_s = f_s(x)$ and $B_s = \Delta_d f_s(x) = \langle w_{s,d}, \psi_s(x) \rangle + \beta_{s,d}$. The discrete product rule $\prod_s (A_s + B_s) - \prod_s A_s = \sum_{\emptyset \neq S' \subseteq [S]} \left(\prod_{s \in S'} B_s \right) \left(\prod_{t \notin S'} A_t \right)$ expresses the gain as a linear form in $\bigotimes_s [\psi_s; f_s; 1]$. $\square$

PROPOSITION 5 (RATIOS: AFFINE COMPARATOR; DA ONLY IN SPECIAL CASES). *Let $F(x) := P(x)/Q(x)$ with $Q(x) > 0$ on P , where P, Q are built from $\{f_s\}$ by finite sums/products. Let Ψ_P, Ψ_Q be their product-lifts and set $\Psi_{\text{rat}}(x) := \Psi_P(x) \otimes \Psi_Q(x)$.*

Then for every EFD d there exists $W_d^{(\text{rat})}$ such that

$$\Delta_d F(x) = \frac{P(x+d)Q(x) - P(x)Q(x+d)}{Q(x)Q(x+d)} = \frac{\langle W_d^{(\text{rat})}, \Psi_{\text{rat}}(x) \rangle}{Q(x)Q(x+d)}.$$

Consequently,

$$\text{sign}(\Delta_d F(x)) = \text{sign}\left(\langle W_d^{(\text{rat})}, \Psi_{\text{rat}}(x) \rangle\right).$$

In general $F \notin \mathcal{DA}^r$ (the denominator depends on x); $F \in \mathcal{DA}^r$ only under exceptional conditions, e.g., when for each d the product $Q(x)Q(x+d)$ is independent of x (in particular, if Q is constant on P).

PROOF. Expand the numerator via Proposition 4 at x and $x+d$ to obtain a linear form in $\Psi_{\text{rat}}(x)$. Since $Q(x)Q(x+d) > 0$, the sign equality follows. Unless $Q(x)Q(x+d)$ is x -independent for each d , the gain cannot be written as an affine function of a finite-dimensional lift, so $F \notin \mathcal{DA}^r$ in general. $\square$

This subsumes linear-fractional [11], polynomial-fractional [17], and finite-product energies [18, 30], although full DA holds only in the special cases noted above.

4.4 Closure under composition

Let $h_j \in \mathcal{DA}^{r_j}$ with feature maps ψ_j so that for every EFD d , $\Delta_d h_j(x) = \langle W_{j,d}, \psi_j(x) \rangle$.

PROPOSITION 6 (AFFINE AND POLYNOMIAL OUTER MAPS PRESERVE DA). (a) If $f(x) = g(h_1(x), \dots, h_k(x))$ with $g(z) = \alpha_0 + \sum_{j=1}^k \alpha_j z_j$ affine, then with

$$\Psi_{\oplus}(x) := \bigoplus_{j=1}^k \psi_j(x) \in \mathbb{R}^{\sum_j r_j}$$

we have $\Delta_d f(x) = \langle W_d^{(\text{aff})}, \Psi_{\oplus}(x) \rangle$ and $f \in \mathcal{DA}^{\sum_j r_j}$.

(b) If $f(x) = g(h_1(x), \dots, h_k(x))$ with g a polynomial of total degree $D \geq 1$, let $H(x) := (h_1(x), \dots, h_k(x))$ and let $\Phi_{\leq D-1}(x)$ be the vector of all monomials in $H(x)$ of total degree $\leq D-1$ (dimension $M = \binom{k+D-1}{D-1}$). Then

$$\Delta_d f(x) = \langle W_d^{(\text{poly})}, \Psi_{\oplus}(x) \otimes \Phi_{\leq D-1}(x) \rangle,$$

so $f \in \mathcal{DA}^R$ with $R \leq \left(\sum_{j=1}^k r_j\right) \binom{k+D-1}{D-1}$.

PROOF. (a) $\Delta_d f = \sum_j \alpha_j \Delta_d h_j$, each term linear in ψ_j ; stack them. (b) Expand g into monomials $\prod_j h_j^{\alpha_j}$ and apply the discrete product rule; every term contains exactly one factor $\Delta_d h_{j^*}$ (linear in ψ_{j^*}) times a product of at most $D-1$ remaining h 's, which is a coordinate of $\Phi_{\leq D-1}(x)$. $\square$

This subsumes hierarchical low-rank energies and shallow polynomial networks over DA primitives, including Boltzmann machines [25] and energy-based models [33, 51].

4.5 Additively separable smooth functions

Consider $f(x) = \sum_{i=1}^n g_i(a_i^\top x)$ with $a_i \in \mathbb{R}^n$. Assume each g_i admits a finite basis expansion of size k_i :

$$(\text{poly}) \quad g_i(z) = \sum_{\ell=0}^{k_i} \alpha_{i\ell} z^\ell \quad \text{or} \quad (\text{exp}) \quad g_i(z) = \sum_{\ell=1}^{k_i} \alpha_{i\ell} e^{\beta_{i\ell} z}.$$

PROPOSITION 7 (FINITE-BASIS SEPARABLE SUMS ARE DA). *Let $\mathcal{D} = \{d^{(1)}, \dots, d^{(m)}\}$ be the EFDs. Define $r := \sum_{i=1}^n k_i$ and the feature map*

$$\psi(x) = \begin{cases} \left((a_i^\top x)^\ell \right)_{i, \ell=0, \dots, k_i} & \text{(poly)} \\ \left(e^{\beta_{i\ell} a_i^\top x} \right)_{i, \ell=1, \dots, k_i} & \text{(exp)} \end{cases} \in \mathbb{R}^{r + (\text{const})}.$$

Then for every $d \in \mathcal{D}$ there exists $w_d \in \left(\mathbb{R}^{r + (\text{const})} \right)^$ such that $\Delta_d f(x) = \langle w_d, \psi(x) \rangle$. Hence $f \in \mathcal{DA}^r$.*

PROOF. (poly) Let $s_i(x) := a_i^\top x$. Then $s_i(x+d)^\ell - s_i(x)^\ell = \sum_{t=0}^{\ell-1} \binom{\ell}{t} s_i(x)^t (a_i^\top d)^{\ell-t}$, linear in $\{s_i(x)^t\}_{t \leq \ell-1}$. Summing over (i, ℓ) yields the claim. (exp) $e^{\beta_{i\ell}(s_i(x)+a_i^\top d)} - e^{\beta_{i\ell}s_i(x)} = \left(e^{\beta_{i\ell}a_i^\top d} - 1 \right) e^{\beta_{i\ell}s_i(x)}$, linear in the exponential feature; summing over (i, ℓ) yields the claim. $\square$

This subsumes finite polynomial truncations and Taylor series approximations [48], mixtures of exponentials [42], finite-basis generalized linear model energies [41].

4.6 Bilinear and multilinear forms

PROPOSITION 8 (LOW-RANK BILINEAR AND MULTILINEAR FORMS ARE DA). **(Bilinear)** *Let $x \in \{0, 1\}^n$, $y \in \{0, 1\}^m$, and $M = \sum_{\ell=1}^r a_\ell b_\ell^\top$ with $a_\ell \in \mathbb{R}^n$, $b_\ell \in \mathbb{R}^m$. Consider*

$$f(x, y) = x^\top M y = \sum_{\ell=1}^r (a_\ell^\top x) (b_\ell^\top y).$$

Define the feature map

$$\psi(x, y) = ((a_\ell^\top x)_{\ell=1}^r, (b_\ell^\top y)_{\ell=1}^r) \in \mathbb{R}^{2r}.$$

For any EFD $d = (d_x, d_y)$, set

$$v_d := ((a_\ell^\top d_x)_{\ell=1}^r, (b_\ell^\top d_y)_{\ell=1}^r) \in (\mathbb{R}^{2r})^*, \quad \beta_d := \sum_{\ell=1}^r (a_\ell^\top d_x) (b_\ell^\top d_y).$$

Then

$$\Delta_d f(x, y) = \langle v_d, \psi(x, y) \rangle + \beta_d,$$

hence $f \in \mathcal{DA}^{2r}$. If EFDs affect only one block at a time (flip in x or in y), then $\beta_d = 0$.

(Multilinear) *Let $t \geq 2$, variables $x^{(s)} \in \{0, 1\}^{n_s}$, and a CP-rank- r decomposition*

$$f(x^{(1)}, \dots, x^{(t)}) = \sum_{\ell=1}^r \prod_{s=1}^t \langle a_\ell^{(s)}, x^{(s)} \rangle, \quad a_\ell^{(s)} \in \mathbb{R}^{n_s}.$$

For each nonempty proper $T \subsetneq [t]$, let

$$\phi_{\ell, T}(x^{(1)}, \dots, x^{(t)}) := \prod_{s \in T} \langle a_\ell^{(s)}, x^{(s)} \rangle,$$

and let ψ collect all $\phi_{\ell, T}$ over $\ell \in [r]$ and $T \subsetneq [t]$, $T \neq \emptyset$. Then for each EFD $d = (d^{(1)}, \dots, d^{(t)})$ there exist v_d, β_d with

$$\Delta_d f = \langle v_d, \psi \rangle + \beta_d.$$

Since there are $2^t - 2$ nonempty proper subsets T and r CP terms, $f \in \mathcal{DA}^{r(2^t-2)}$. If EFDs affect at most one block $x^{(s)}$, it suffices to keep only $T = [t] \setminus \{s\}$, giving $f \in \mathcal{DA}^{rt}$.

PROOF. **Bilinear:** With $f = \sum_{\ell=1}^r (a_\ell^\top x)(b_\ell^\top y)$ and $d = (d_x, d_y)$,

$$\Delta_d f = \sum_{\ell=1}^r [(a_\ell^\top d_x)(b_\ell^\top y) + (a_\ell^\top x)(b_\ell^\top d_y) + (a_\ell^\top d_x)(b_\ell^\top d_y)] = \langle v_d, \psi(x, y) \rangle + \beta_d.$$

Multilinear: Write $u_\ell^{(s)}(x) := \langle a_\ell^{(s)}, x^{(s)} \rangle$ and $\delta_\ell^{(s)} := \langle a_\ell^{(s)}, d^{(s)} \rangle$. Then

$$\Delta_d \left(\prod_{s=1}^t u_\ell^{(s)} \right) = \prod_{s=1}^t (u_\ell^{(s)} + \delta_\ell^{(s)}) - \prod_{s=1}^t u_\ell^{(s)} = \sum_{\emptyset \neq S \subseteq [t]} \left(\prod_{s \in S} \delta_\ell^{(s)} \right) \left(\prod_{s \notin S} u_\ell^{(s)} \right),$$

an affine combination of partial products $\phi_{\ell,T}$ with $T = [t] \setminus S$ (the $S = [t]$ term yields β_d). Summing over ℓ yields the DA decomposition and the stated dimension bounds. $\square$

This subsumes low-rank assignment [32] and matrix factorization energies [34], and CP-rank- r multilinear objectives [26].

5 Applications

5.1 Trimmed and robust statistics

See [47] for an introduction to robust statistics.

REMARK (DIFFERENTIALLY LINEAR RELAXATION). We write $f \in \mathcal{DL}^r$ (differentially linear) if discrete gains admit $\Delta_d f(x) = \langle w_d, \psi(x) \rangle / \Pi_d(x)$ where $\Pi_d(x) > 0$. Since signs are preserved, $\mathcal{DL}$ suffices for chamber enumeration. Note $\mathcal{DA}^r \subset \mathcal{DL}^r$.

Let $X \in \mathbb{R}^{n \times d}$, $y \in \mathbb{R}^n$, and $x \in \{0, 1\}^n$ select a subset via $W(x) = \text{Diag}(x)$. Define the linear aggregates

$$k = \mathbf{1}^\top x, \quad T_{rs}(x) = \sum_i x_i X_{ir} X_{is}, \quad s_r(x) = \sum_i x_i X_{ir} y_i, \quad S_{yy}(x) = \sum_i x_i y_i^2,$$

and collect $A(x) := \{k, S_{yy}, T_{rs}, s_r\}$ with $K := |A| = \frac{d(d+1)}{2} + d + 2$.

PROPOSITION 9 (RATIONAL-IN-AGGREGATES ARE $\mathcal{DL}$). Let $F(x) = P(A(x))/Q(A(x))$ where P, Q are polynomials of total degrees D_P, D_Q in the entries of $A(x)$, and assume $Q(A(x)) > 0$ on the feasible region and $\text{rank} T(x) = d$ when T^{-1} appears. Set $D := D_P + D_Q - 1$ and define the Veronese feature map [40] $\psi : \{0, 1\}^n \rightarrow \mathbb{R}^R$ stacking all monomials in $A(x)$ of total degree $\leq D$, with $R = \binom{K+D}{D}$. Then for every elementary feasible direction d there exist $w_d \in (\mathbb{R}^R)^*$ and a positive $\Pi_d(x)$ such that

$$\Delta_d F(x) = \frac{\langle w_d, \psi(x) \rangle}{\Pi_d(x)}.$$

Hence $F \in \mathcal{DL}^R$, and exact optimization over any feasible family with a linear-optimization oracle runs in $O(m^r)$ by Theorem 2.

REMARK. $\mathcal{DL}^R$ is the class of functions whose discrete gains admit an affine comparator up to a positive scaling factor, so $\mathcal{DA}^R \subset \mathcal{DL}^R$. This suffices for sign-consistent chamber enumeration.

COROLLARY 2 (TRIMMED COVARIANCE). For $F(x) = \text{Cov}(X^\top x, y^\top x)$ under fixed subset size $k = \mathbf{1}^\top x$, F is linear-fractional in $A(x)$ with $(D_P, D_Q) = (2, 2)$, giving $D = 3$ and $R = \binom{K+3}{3}$; thus $F \in \mathcal{DL}^R$ and is optimizable in $O(T(n, m)m^{\binom{K+3}{3}})$ time.

COROLLARY 3 (TRIMMED R^2). For one-dimensional y and d -dimensional X , $F(x) = R^2(x) = \text{Cov}^2 / (\text{Var}_x \text{Var}_y)$ is a ratio with $(D_P, D_Q) = (4, 2)$, hence $D = 5$ and $R = \binom{K+5}{5}$. Therefore $R^2 \in \mathcal{DL}^R$ and admits $O\left(m^{\binom{K+5}{5}}\right)$ optimization.

COROLLARY 4 (TRIMMED BLUE SLOPES). For $\hat{\beta}(x) = (X^\top W X)^{-1} X^\top W y$, each coordinate $\hat{\beta}_j(x) = P_j(A(x))/Q_j(A(x))$ with $D_P \leq d+1$, $D_Q \leq d$, so $D \leq 2d$ and $R = \binom{K+2d}{2d}$; thus $\hat{\beta}_j \in \mathcal{DL}^R$.

COROLLARY 5 (TRIMMED GAUSS-MARKOV RSS). $\text{RSS}(x) = S_{yy}(x) - s(x)^\top T(x)^{-1} s(x)$ equals $P(A(x))/Q(A(x))$ with $D_P \leq d+1$, $D_Q \leq d$, hence $D \leq 2d$ and $R = \binom{K+2d}{2d}$; therefore $\text{RSS} \in \mathcal{DL}^R$.

5.2 Optimal experimental design

DA also applies to D-, A-, and E-optimality in optimal experimental design (OED) [28]. Let $a_i \in \mathbb{R}^d$, $x \in \{0, 1\}^n$, and $M(x) = \sum_{i=1}^n x_i a_i a_i^\top \in \mathbb{R}^{d \times d}$. Write the quadratic Veronese features [40] $\phi(x) = \text{vecs}(M(x)) \in \mathbb{R}^K$ with $K = d(d+1)/2$. For an integer $D \geq 0$, let $\psi_D(x)$ collect all monomials in the coordinates of $\phi(x)$ of total degree $\leq D$ (including the constant 1).

PROPOSITION 10 (DIFFERENTIAL LINEARITY OF CANONICAL OED CRITERIA). The following objectives are EFD-wise differentially linear with respect to suitable lifts ψ_D : (i) D-optimality $f_D(x) = \det M(x)$ with $D = d-1$; (ii) A-optimality $f_A(x) = -\text{tr}(M(x)^{-1}) = -\frac{\text{adj}(M(x)) : I}{\det M(x)}$ with $D = 2d-2$; (iii) for any fixed $\tau \in \mathbb{R}$, the E-optimality decision polynomial $g_\tau(x) = \det(M(x) - \tau I)$ with $D = d-1$. Consequently, there exist covectors W_d (one per elementary feasible direction d) such that for all feasible x and EFDs d ,

$$\Delta_d f_D(x) = \left\langle W_d^{(D)}, \psi_{d-1}(x) \right\rangle, \quad \Delta_d f_A(x) = \frac{\left\langle W_d^{(A)}, \psi_{2d-2}(x) \right\rangle}{\det M(x) \det M(x+d)}, \quad \Delta_d g_\tau(x) = \left\langle W_d^{(\tau)}, \psi_{d-1}(x) \right\rangle.$$

In particular, $f_D \in \mathcal{DL}^{R_D}$, $f_A \in \mathcal{DL}^{R_A}$, and $g_\tau \in \mathcal{DL}^{R_E}$ with

$$R_D = \binom{K+d-1}{d-1}, \quad R_A = \binom{K+2d-2}{2d-2}, \quad R_E = \binom{K+d-1}{d-1}.$$

PROOF SKETCH. Each entry of M is linear in x , so $\det M$ is a degree- d polynomial in $\phi(x)$; for any EFD d , the finite difference $\det(M(x+d)) - \det(M(x))$ is a polynomial in $\phi(x)$ of degree $\leq d-1$, hence a linear form in $\psi_{d-1}(x)$. For A-optimality, $\text{adj}(M)$ and $\det(M)$ are degree $d-1$ and d polynomials in $\phi(x)$; clearing denominators in $\Delta_d f_A$ yields a numerator polynomial of degree $\leq (d-1) + (d-1) = 2d-2$, i.e., a linear form in $\psi_{2d-2}(x)$, divided by the positive factor $\det M(x) \det M(x+d)$. For E-optimality in decision form, $g_\tau(x) = \det(M(x) - \tau I)$ is a degree- d polynomial in $\phi(x)$, so its finite difference is linear in $\psi_{d-1}(x)$. $\square$

5.3 Portfolio optimization

Classical mean-variance portfolio selection [39] with factor risk [20, 49] admits a $\mathcal{DL}$ reduction under equal-weighted subset selection. Let assets $i \in [n]$ have expected returns μ_i , idiosyncratic volatilities $\sigma_i > 0$, and p -factor loadings $b_{i\ell}$ with diagonal factor covariance $\Lambda = \text{Diag}(\lambda_1, \dots, \lambda_p) \succ 0$; the covariance is $\Sigma = B\Lambda B^\top + \text{Diag}(\sigma_1^2, \dots, \sigma_n^2)$. For a selection $x \in \{0, 1\}^n$ of size $k = \mathbf{1}^\top x$, the equal-weight portfolio has mean

$$\mu(x) = \frac{1}{k} \sum_i \mu_i x_i \quad \text{and} \quad \text{Var}(x) = \frac{1}{k^2} \left(\sum_{\ell=1}^p \lambda_\ell F_\ell(x)^2 + D_{\text{id}}(x) \right),$$

where the *portfolio aggregates* are linear in x :

$$U(x) := \sum_i \mu_i x_i, \quad F_\ell(x) := \sum_i b_{i\ell} x_i \quad (\ell \in [p]), \quad S_\sigma(x) := \sum_i \sigma_i x_i, \quad D_{\text{id}}(x) := \sum_i \sigma_i^2 x_i.$$

Collect $A(x) := \{k, U, F_1, \dots, F_p, S_\sigma, D_{\text{id}}\}$ with $K = p + 4$ (all linear in x).

PROPOSITION 11 (RATIONAL-IN-PORTFOLIO-AGGREGATES ARE $\mathcal{DL}$). *Let $F(x) = P(A(x))/Q(A(x))$ with P, Q polynomials of total degrees D_P, D_Q in the entries of $A(x)$, and assume $Q(A(x)) > 0$ on the feasible region (true here since $k \geq 1$ and variances are positive). Set $D := D_P + D_Q - 1$ and let $\psi : \{0, 1\}^n \rightarrow \mathbb{R}^R$ be the Veronese lift stacking all monomials in $A(x)$ of total degree $\leq D$ [40], with $R = \binom{K+D}{D}$. Then for every EFD d there exist $w_d \in (\mathbb{R}^R)^*$ and a positive $\Pi_d(x)$ such that*

$$\Delta_d F(x) = \frac{\langle w_d, \psi(x) \rangle}{\Pi_d(x)}.$$

Hence $F \in \mathcal{DL}^R$ and exact optimization over any feasible family with a linear-optimization oracle runs in $O(m^R)$ by Theorem 2.

COROLLARY 6 (MEAN-VARIANCE TRADEOFF). *For $F_{\text{MV}}(x) = \alpha\mu(x) - \beta\text{Var}(x)$ with $\alpha, \beta \geq 0$, clearing denominators gives $F_{\text{MV}}(x) = \frac{\alpha k U - \beta(\sum_\ell \lambda_\ell F_\ell^2 + D_{\text{id}})}{k^2}$, so $(D_P, D_Q) = (2, 2)$ and $D = 3$. Thus $F_{\text{MV}} \in \mathcal{DL}^R$ with $R = \binom{K+3}{3}$ and admits $O(m^{\binom{K+3}{3}})$ optimization.*

COROLLARY 7 (SQUARED SHARPE RATIO). *The squared Sharpe ratio of the equal-weight portfolio is $S^2(x) = \mu(x)^2 / \text{Var}(x) = \frac{U^2}{\sum_\ell \lambda_\ell F_\ell^2 + D_{\text{id}}}$, a ratio with $(D_P, D_Q) = (2, 2)$, hence $D = 3$ and $R = \binom{K+3}{3}$. Therefore $S^2 \in \mathcal{DL}^R$ with the same complexity bound.*

COROLLARY 8 (DIVERSIFICATION RATIO (SQUARED)). *The diversification ratio [12, 38] for equal weights is $\text{DR}(x) = \frac{\sum_i w_i \sigma_i}{\sqrt{w^\top \Sigma w}} = \frac{S_\sigma/k}{\sqrt{\text{Var}(x)}}$. Then $\text{DR}^2(x) = \frac{S_\sigma^2}{\sum_\ell \lambda_\ell F_\ell^2 + D_{\text{id}}}$ has $(D_P, D_Q) = (2, 2)$, so $D = 3$, $R = \binom{K+3}{3}$, and $\text{DR}^2 \in \mathcal{DL}^R$.*

6 Relation to existing paradigms

Our rank parameter serves as a structural width measure: bounding it yields fixed-parameter or XP-time tractability, analogous to treewidth in Courcelle-type meta-theorems [19] and to Grohe's structural parameters for constraint satisfaction [22]. In contrast to low-rank continuous relaxations for nonlinear programs, our approach is purely combinatorial, operating directly on the discrete hyperplane arrangement and producing exact solutions.

Prior to our work, even for QUBO, only limited classes of nonlinear objectives were known to be tractable, e.g., where Q is positive semidefinite [8], a sum matrix, a product matrix (particularly in the quadratic assignment problem [10]), a Toeplitz matrix [9], has non-negative off-diagonal elements [35, 43, 44], etc. Other polynomial solutions were limited to cases where Q represents a tractable graph structure like a planar graph [36, 37], graph of bounded bandwidth [16], graph of bounded treewidth [4, 5], graph of bounded genus [21], perfect graph [24], t -perfect graphs [23], interaction graphs, [7], etc. For Ising models, some prior work had explored separable energies [2, 6]. See [46] for a complete list.

Related work by Del-Pia and Khajavirad [13–15] gives polynomial-size extended formulations for binary polynomials with few multiplicative factors solved via mixed-integer programming. Our framework is complementary and more general in key respects, namely in subsuming such factorized models via suitable ψ without convexification and yields direct combinatorial enumeration in $n^{O(r)}$ time. Solving polynomial programs by extending methods from quadratic programs had already been considered [1], albeit without the same XP guarantees. Dinklebach's method [17] and Charnes-Cooper transform [11] drew a connection between fractional programs and established methods in polynomial and linear programming, with guarantees of tractability in the latter case.

7 Conclusion and outlook

We introduced *differential affineness* (DA) as a unifying structural parameter for nonlinear pseudo-Boolean optimization and proved an XP metatheorem: whenever an objective f has DA rank r , maximizing f reduces to enumerating $O(m^r)$ chambers of an explicit affine hyperplane arrangement (where m is the number of elementary directions), together with a two-sided local-optimality test yielding $O(n)$ -time decoding per chamber (Theorem 2). The framework is constructive: we provided closed-form lifts ψ , covectors $\{v_d\}$, and offsets $\{\beta_d\}$ for broad objective classes—low-rank quadratics, low-Waring-rank polynomials, products, compositions, separable sums, and CP-rank multilinear forms—and instantiated applications in robust statistics, experimental design, and portfolio optimization. Conceptually, DA rank plays a Courcelle-style width role for nonlinear objectives, replacing logical width (treewidth) with an analytic rank governing discrete gains.

7.1 Limitations

- (1) *Ratios and DL structure.* Fractional objectives P/Q generally require directional comparisons when Q varies; full DA holds only when Q is constant along direction pairs.
- (2) *Rank blowups.* Composition and polynomial lifts can increase r combinatorially (Veronese growth), preserving XP status but affecting practical constants.
- (3) *Constrained multiplicity.* With compound moves (e.g., swaps $\{\pm(e_i - e_j)\}$ for ranking), chambers may admit multiple stable points; uniqueness requires two-sided sign opposition at every direction.
- (4) *Model extraction.* Obtaining Waring or CP decompositions may be costly or approximate; our guarantees then apply to the extracted surrogate.

7.2 Open problems

- (1) *Constrained domains.* Extend to budget, knapsack, matroid, and exchange polytopes by characterizing $|\mathcal{D}|$ and exploiting arrangement sparsity.
- (2) *Tighter rank bounds.* Close gaps between worst-case Veronese dimensions and problem-specific structure (symmetries, separability). Characterize minimal DA rank and its identifiability.
- (3) *Tope-graph algorithms.* Replace full enumeration with output-sensitive walks exploiting oriented-matroid pivots, Gray-code sign updates, or incremental feature evaluation.
- (4) *Learning ψ .* Given oracle access to $\Delta_d f(x)$, recover a minimal lift via structured low-rank regression with sample-complexity guarantees.
- (5) *Smoothed analysis.* Prove that under random perturbations of $\{v_d, \beta_d\}$, opposite-sign conditions hold with high probability and the number of relevant chambers is polynomial.
- (6) *Lower bounds.* Establish $W[1]$ /ETH-style barriers when r grows, or hardness when ψ is unknown or only comparison oracles are available.

DA rank isolates the combinatorial essence of nonlinear objectives: when discrete gains lie in a low-dimensional affine model, exact optimization becomes chamber enumeration in that dimension. This analytic width measure complements classical structural parameters (treewidth, rank-width) and provides a tractable, explicit route through problems previously requiring heavy relaxations. The framework’s practical impact will grow with tighter rank bounds, extensions to constrained domains, and efficient tope-walk solvers.

References

- [1] Martin Anthony, Endre Boros, Yves Crama, and Aritanan Gruber. 2017. Quadratic reformulations of nonlinear binary optimization problems. *Mathematical Programming* 162, 1 (2017), 115–144.
- [2] Francisco Barahona. 1982. On the computational complexity of Ising spin glass models. *Journal of Physics A: Mathematical and General* 15, 10 (1982), 3241–3253. doi:10.1088/0305-4470/15/10/028
- [3] Géraud Blatman and Bruno Sudret. 2011. Adaptive sparse polynomial chaos expansion based on least angle regression. *Journal of computational Physics* 230, 6 (2011), 2345–2367.
- [4] Hans L Bodlaender and Klaus Jansen. 2000. On the complexity of the maximum cut problem. *Nordic Journal of Computing* 7, 1 (2000), 14–31.
- [5] Hans L Bodlaender and Arie MCA Koster. 2008. Combinatorial optimization on graphs of bounded treewidth. *Comput. J.* 51, 3 (2008), 255–269.
- [6] Endre Boros and Peter L. Hammer. 2002. *Pseudo-Boolean Optimization*. Springer. doi:10.1007/978-1-4613-0017-8
- [7] Endre Boros, Peter L. Hammer, and Rong Sun. 2007. Pseudo-Boolean optimization: A survey. *Discrete Applied Mathematics* 155, 4 (2007), 495–539. doi:10.1016/j.dam.2006.06.010
- [8] Nicolas Boumal, Vlad Voroninski, and Afonso S. Bandeira. 2016. The non-convex Burer–Monteiro approach works on smooth semidefinite programs. In *Advances in Neural Information Processing Systems 29 (NeurIPS 2016)*.
- [9] Rainer E Burkard, Eranda Cela, Günter Rote, and Gerhard J Woeginger. 1998. The quadratic assignment problem with a monotone anti-Monge and a symmetric Toeplitz matrix: Easy and hard cases. *Mathematical Programming* 82, 1 (1998), 125–158.
- [10] Eranda Cela, Vladimir G Deineko, and Gerhard J Woeginger. 2015. Well-solvable cases of the QAP with block-structured matrices. *Discrete applied mathematics* 186 (2015), 56–65.
- [11] Abraham Charnes and William W Cooper. 1962. Programming with linear fractional functionals. *Naval Research logistics quarterly* 9, 3-4 (1962), 181–186.
- [12] Yves Choueifaty. 2008. Towards maximum diversification. *Available at SSRN 4063676* (2008).
- [13] Alberto Del Pia. 2025. Factorized Binary Polynomial Optimization. *Mathematical Programming* (2025). doi:10.1007/s10107-025-02274-5
- [14] Alberto Del Pia and Abhishek Khajavirad. 2017. A Polyhedral Study of Binary Polynomial Programs. *Mathematics of Operations Research* 42, 2 (2017), 389–410. doi:10.1287/moor.2016.0804
- [15] Alberto Del Pia and Abhishek Khajavirad. 2023. The pseudo-Boolean polytope and polynomial-size extended formulations for binary polynomial optimization. *arXiv preprint arXiv:2309.08693* (2023). <https://arxiv.org/abs/2309.08693>
- [16] Mauro Dell’Amico, Francesco Maffioli, Silvano Martello, et al. 1997. *Annotated bibliographies in combinatorial optimization*. Wiley.
- [17] Werner Dinkelbach. 1967. On nonlinear fractional programming. *Management science* 13, 7 (1967), 492–498.
- [18] Joseph G Ecker. 1980. Geometric programming: methods, computations and applications. *SIAM review* 22, 3 (1980), 338–362.
- [19] Michael Elberfeld et al. 2016. Variants of Courcelle’s Theorem for complexity classes. *ACM Transactions on Algorithms* 12, 3 (2016), 1–19.
- [20] Eugene F Fama and Kenneth R French. 1992. The cross-section of expected stock returns. *the Journal of Finance* 47, 2 (1992), 427–465.
- [21] Anna Galluccio, Martin Loeb, and Jan Vondrák. 2001. Optimization via enumeration: a new algorithm for the max cut problem. *Mathematical Programming* 90, 2 (2001), 273–290.
- [22] Martin Grohe. 2007. The complexity of homomorphism and constraint satisfaction problems seen from the other side. *Journal of the ACM (JACM)* 54, 1 (2007), 1–24.
- [23] Martin Grötschel and László Lovász. 1995. Combinatorial optimization. *Handbook of combinatorics* 2, 1541-1597 (1995), 4.
- [24] Martin Grötschel, László Lovász, and Alexander Schrijver. 1984. Polynomial algorithms for perfect graphs. In *North-Holland mathematics studies*. Vol. 88. Elsevier, 325–356.
- [25] Geoffrey E Hinton, Terrence J Sejnowski, et al. 1986. Learning and relearning in Boltzmann machines. *Parallel distributed processing: Explorations in the microstructure of cognition* 1, 282-317 (1986), 2.
- [26] Frank L Hitchcock. 1927. The expression of a tensor or a polyadic as a sum of products. *Journal of Mathematics and Physics* 6, 1-4 (1927), 164–189.
- [27] Ernst Ising. 1925. Beitrag zur theorie des ferromagnetismus. *Zeitschrift für Physik* 31, 1 (1925), 253–258.
- [28] Jack Kiefer. 1959. Optimum experimental designs. *Journal of the Royal Statistical Society: Series B (Methodological)* 21, 2 (1959), 272–304.
- [29] Tamara G Kolda and Brett W Bader. 2009. Tensor decompositions and applications. *SIAM review* 51, 3 (2009), 455–500.
- [30] Daphne Koller and Nir Friedman. 2009. *Probabilistic graphical models: principles and techniques*. MIT press.
- [31] John Lafferty, Andrew McCallum, and Fernando CN Pereira. 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. (2001).
- [32] Eugene L Lawler. 1963. The quadratic assignment problem. *Management science* 9, 4 (1963), 586–599.
- [33] Yann LeCun, Sumit Chopra, Raia Hadsell, M Ranzato, Fuyie Huang, et al. 2006. A tutorial on energy-based learning. *Predicting structured data* 1, 0 (2006).
- [34] Daniel D Lee and H Sebastian Seung. 1999. Learning the parts of objects by non-negative matrix factorization. *nature* 401, 6755 (1999), 788–791.
- [35] Duan Li, Xiaoling Sun, Shenshen Gu, Jianjun Gao, and Chunli Liu. 2010. Polynomially solvable cases of binary quadratic programs. In *Optimization and Optimal Control: Theory and Applications*. Springer, 199–225.
- [36] Frauke Liers and Gregor Pardella. 2012. Partitioning planar graphs: a fast combinatorial approach for max-cut. *Computational Optimization and Applications* 51, 1 (2012), 323–344.

- [37] Richard J Lipton and Robert Endre Tarjan. 1979. A separator theorem for planar graphs. *SIAM J. Appl. Math.* 36, 2 (1979), 177–189.
- [38] Sbastien Maillard, Thierry Roncalli, and Jerome Teiletche. 2008. On the properties of equally-weighted risk contributions portfolios. SSRN.
- [39] Harry Markowitz. 1952. Modern portfolio theory. *Journal of Finance* 7, 11 (1952), 77–91.
- [40] Jiri Matousek. 2013. *Lectures on discrete geometry*. Vol. 212. Springer Science & Business Media.
- [41] Peter McCullagh. 2019. *Generalized linear models*. Routledge.
- [42] Geoffrey J McLachlan and David Peel. 2000. *Finite mixture models*. John Wiley & Sons.
- [43] Jean-Claude Picard and Maurice Queyranne. 1982. Selected applications of minimum cuts in networks. *INFOR: Information Systems and Operational Research* 20, 4 (1982), 394–422.
- [44] Jean-Claude Picard and H Donald Ratliff. 1975. Minimum cuts and related problems. *Networks* 5, 4 (1975), 357–370.
- [45] Renfrey Burnard Potts. 1952. Some generalized order-disorder transformations. In *Mathematical proceedings of the cambridge philosophical society*, Vol. 48. Cambridge University Press, 106–109.
- [46] Abraham P Punnen. 2022. The quadratic unconstrained binary optimization problem. *Springer International Publishing* 10 (2022), 978–3.
- [47] Peter J Rousseeuw and Mia Hubert. 2011. Robust statistics for outlier detection. *Wiley interdisciplinary reviews: Data mining and knowledge discovery* 1, 1 (2011), 73–79.
- [48] Walter Rudin. 1987. *Real and complex analysis*. McGraw-Hill, Inc.
- [49] William F Sharpe. 1963. A simplified model for portfolio analysis. *Management science* 9, 2 (1963), 277–293.
- [50] David Sherrington and Scott Kirkpatrick. 1975. Solvable model of a spin-glass. *Physical review letters* 35, 26 (1975), 1792.
- [51] Paul Smolensky. 1986. *Information processing in dynamical systems: Foundations of harmony theory*. Technical Report.
- [52] T. Zaslavsky. 1975. Facing up to arrangements: face-count formulas for partitions of space by hyperplanes. *Memoirs of the American Mathematical Society* 154 (1975).