International Journal of Computing and Optimization
Vol. 13, 2026, no. 1, 1 - 19
HIKARI Ltd, www.m-hikari.com
https://doi.org/10.12988/ijco.2026.10017

Squared-Variable Parametrizations in Inequality-Constrained Optimization: Benign Nonconvexity, Singular Geometry, and Second-Order Transfer

Kan-Lin Hsiung

Department of Electrical Engineering, Yuan-Ze University
Taoyuan City 320315, Taiwan

Abstract

Squared-variable parametrizations encode nonnegativity through algebraic identities such as $x = v^2$ or $c_i(x) - s_i^2 = 0$, thereby replacing inequality constraints by equalities. Their appeal is clear, but so is the cost: the lift is singular at active constraints, the lifted variables acquire sign symmetry, and the reformulated problem is usually nonconvex even when the original problem is convex. This survey argues that these effects are best understood not as implementation details or fatal pathologies, but as a coherent second-order phenomenon. For simplex constraints, the Hadamard lift to the sphere preserves the KKT/strict-saddle dichotomy, so convex simplex problems inherit a benign nonconvex landscape. For polyhedral feasible sets, the lift produces a singular algebraic variety whose tangent cones and second-order tangent sets support multiplier recovery and optimality certification without additional constraint qualifications. For bound-constrained and general nonlinear programs, exact and approximate transfer theorems link second-order properties of direct square substitution and squared-slack formulations to those of the original problem, with quantitative residual bounds governed by thresholded active-set information. The resulting picture is not that squaring is costless, but that its singularities are structured enough to support a rigorous theory of landscape, certification, and algorithmic transfer.

Mathematics Subject Classification: 90C26, 90C30, 90C46

Keywords: squared-variable formulation, Hadamard parametrization, LICQ, algebraic variety, Riemannian optimization, second-order optimality

1 Introduction

1.1 Motivation

Inequality constraints are usually handled directly by active-set, interior-point, or projection-based methods. Squared-variable parametrizations take a different route: they encode nonnegativity in the variables themselves, replacing inequalities by equalities and, in some settings, removing explicit inequality constraints altogether.

The attraction is immediate. A bound such as $x_i \geq 0$ can be enforced exactly by writing $x_i = y_i^2$; more generally, $c_i(x) \geq 0$ can be rewritten as $c_i(x) - s_i^2 = 0$. Both devices are classical; see, for example, [2, 6]. The reformulated problem is then accessible to methods for equality-constrained or even unconstrained optimization. The price is equally immediate. Squaring introduces sign symmetry, destroys convexity in the ambient variables, and becomes singular at the boundary because the map $t \mapsto t^2$ has zero derivative at the origin.

For a long time, these drawbacks were treated as decisive in practice. The reformulations were acknowledged as useful expedients, especially when one wished to reuse equality-constrained software, yet they remained theoretically awkward: first-order stationarity no longer encoded the correct multiplier signs, and classical regularity had to be reconsidered precisely where complementarity became active.

Recent work supports a sharper conclusion. The central issue is not whether squaring is globally benign, but whether the defects introduced by the lift are fatal and, if not, which mathematical objects recover the missing information. Does the lift create harmful critical points? How should one certify optimality when the lifted feasible set is singular? Which exact and approximate second-order guarantees survive the passage from the original formulation to the squared one? These questions organize the survey.

This selective, expository survey focuses on papers that resolve the classical objections not by denying them but by replacing first-order intuition with the appropriate second-order geometry.

Two presentational choices guide the discussion. First, we organize the survey by mathematical obstruction rather than by publication date: benign nonconvexity on a smooth manifold, controlled singularity on a parametrized variety, and second-order transfer for general nonlinear programs. Second, we emphasize what each paper contributes at the level of objects rather than slogans: Riemannian Hessians in the simplex case, tangent cones and second-order tangent sets in the polyhedral case, and thresholded residual transfer in the approximate nonlinear-programming case. This object-level view makes the recent literature fit together as a coherent theory rather than as a sequence of isolated observations.

1.2 Two variants

Two closely related transformations should be kept separate. We adopt the sign convention $c_i(x) \geq 0$; the equivalent convention $g_i(x) \leq 0$ is obtained by setting $c_i = -g_i$.

Direct Square Substitution (DSS). For explicit nonnegativity constraints, consider

$$\min_{x} \; f(x) \quad \text{s.t. } Ax = b, \; x \geq 0. \tag{LCP}$$

Substituting $x_i = y_i^2$, or equivalently $x = y \circ y$, yields

$$\min_{y} \; f(y \circ y) \quad \text{s.t. } A(y \circ y) = b. \tag{LCPH}$$

The nonnegativity constraint is enforced implicitly and disappears from the formulation. This is the setting studied by Li, McKenzie, and Yin [8] and by Tang and Toh [11].

Squared-Slack Variables (SSV). For a problem with general inequality constraints, consider

$$\min_{x} \; f(x) \quad \text{s.t. } c_i(x) \geq 0, \; i = 1, \ldots, m. \tag{NLP}$$

Introducing slack variables s_i yields

$$\min_{x,s} \; f(x) \quad \text{s.t. } c_i(x) - s_i^2 = 0, \; i = 1, \ldots, m. \tag{SSV}$$

This is the form studied by Ding and Wright [3] and, in an earlier note, by Fukuda and Fukushima [5].

The common algebraic feature is the squaring map $t \mapsto t^2$. Because this map enforces nonnegativity automatically but has derivative equal to zero at $t = 0$, boundary points of the original feasible set become singular points of the reformulated equality system.

The distinction between DSS and SSV matters throughout the survey. The simplex and polyhedral results concern DSS, where the lift changes the geometry of the feasible set itself. The general nonlinear-programming results concern SSV, where the objective is left untouched but the reformulated equality system becomes singular at active inequalities. The two branches share the same algebraic mechanism but require different analytical tools.

1.3 Stationarity language and scope

We use standard KKT terminology for the original inequality-constrained problem, with multipliers chosen according to the convention $c_i(x) \geq 0$. Thus one may write the Lagrangian as $f(x) - \lambda^T c(x)$, and the inequality multipliers satisfy $\lambda_i \geq 0$. The active set at a feasible point is denoted $\mathcal{A}(x) = \{i : c_i(x) = 0\}$. For the squared-slack formulation, the equality multipliers are unrestricted in sign; this is why first-order stationarity of the lifted system is not, by itself, equivalent to KKT stationarity of the original problem.

The phrase *weak second-order necessary condition* is used in the sense common in the recent squared-variable literature: curvature is controlled on the linearized critical directions naturally recovered from the reformulation, rather than on every directional set appearing in stronger copositivity-based second-order conditions. This distinction matters because several exact transfer results yield precisely weak second-order necessary conditions. Accordingly, the theorem statements below are informal syntheses; the cited papers provide the fully quantified constraint qualifications, residual definitions, and active-set thresholds.

We do not cover every use of slack variables, every equality-constrained algorithm, or every reparametrization of nonnegative variables. Instead, we focus on recent work that analyzes the singularity of the squaring map explicitly and uses that analysis to recover optimality information.

1.4 A running example

A simple example will guide the discussion. Consider the linear program

$$\min\ -x_1 - 2x_2 \quad \text{s.t. } x_1 + x_2 = 1,\ x_1 \geq 0,\ x_2 \geq 0. \tag{Ex}$$

The feasible set is the line segment $\{(x_1, x_2) : x_1 + x_2 = 1,\ x_1, x_2 \geq 0\}$, namely the one-dimensional probability simplex. The unique optimal solution is $x^* = (0, 1)$, where the bound $x_1 \geq 0$ is active.

Under DSS, we set $x_i = y_i^2$ and obtain

$$\min\ -y_1^2 - 2y_2^2 \quad \text{s.t. } y_1^2 + y_2^2 = 1. \tag{Ex-H}$$

This is an optimization problem on the unit circle S^1. The two optimal lifts are $y^* = (0, \pm 1)$, both mapping to the same primal solution $x^* = (0, 1)$. We return to this example whenever the general theory benefits from a concrete picture.

1.5 Outline

The survey proceeds from the cleanest setting to the most general one. Section 2 states the classical objections in a form sharp enough to motivate the later theory. Section 3 treats the simplex-to-sphere lift, where the principal issue is whether the induced nonconvexity creates genuinely harmful critical points. Section 4 turns to polyhedral feasible sets, where the lift is singular and tangent-cone geometry becomes the right language. Section 5 discusses second-order transfer results for bound-constrained and general nonlinear programs. Section 6 synthesizes the common mechanism, relates the squared-variable story to broader theories of smooth parametrization, and highlights limitations and open directions. Section 7 concludes the survey.

2 Classical Objections

The classical objections to squared-variable formulations are substantive. Each identifies a genuine change produced by the squaring map, and each reappears later in a more precise form. We therefore begin by stating these objections sharply enough to clarify what the recent theory resolves.

2.1 Loss of convexity

A squared-variable reformulation may be nonconvex even when f and the constraint functions c_i are convex. In the SSV setting, the reformulation (SSV) is nonconvex in the joint (x, s) space because the relation $c_i(x) = s_i^2$ is quadratic rather than affine. In the DSS setting, the feasible set may become a smooth manifold, whereas the lifted objective is typically nonconvex in the new variables. Although the sources of nonconvexity differ, the basic guarantee of convex optimization—that every local minimizer is global—no longer holds after reformulation.

Example 2.1 (Running example)**.** The objective of (Ex-H), namely $g(y) = -y_1^2 - 2y_2^2$, is concave as a function of y, even though the original problem (Ex) is linear. The constraint $y_1^2 + y_2^2 = 1$ is nonconvex as a subset of $\mathbb{R}^2$. Nevertheless, the resulting nonconvex problem has a benign critical-point structure.

2.2 Boundary degeneracy and regularity

For the SSV reformulation, define $h_i(x, s) = c_i(x) - s_i^2$. Its derivative with respect to s_i is

$$\frac{\partial h_i}{\partial s_i} = -2s_i,$$

which vanishes whenever the ith original inequality is active ($c_i(x^*) = 0$ implies $s_i^* = 0$). Hence

$$\nabla h_i(x^*, s^*) = \begin{pmatrix} \nabla_x c_i(x^*) \\ -2s_i^* \end{pmatrix},$$

so the diagonal slack block that would be nonsingular away from the boundary disappears precisely on the active set.

The right conclusion is subtler than the common slogan that the linear independence constraint qualification (LICQ) "fails after squaring." Ding and Wright show that, at corresponding feasible points, LICQ for the original problem and LICQ for the squared-slack reformulation are equivalent [3]. What vanishes are the slack columns attached to active constraints; whether full row rank is lost depends on the active primal gradients. The genuine difficulty is therefore not automatic rank failure but a change in how active-set information is encoded exactly where complementarity matters.

Example 2.2 (Running example)**.** At $y^* = (0, 1)$ for (Ex-H), the unique equality constraint has gradient $\nabla(y_1^2 + y_2^2) = (0, 2)$, so LICQ holds. If, instead, one introduced a separate squared-slack variable for $x_1 \geq 0$, the associated slack derivative would vanish at the optimum; yet LICQ could still hold because the active primal gradient remains nonzero. The lesson is not automatic rank failure but rather that the slack variables cease to carry first-order information at the boundary. This illustrates a practical difference between DSS and SSV: the former may avoid some of the boundary degeneracy that the latter makes explicit.

2.3 Spurious first-order stationary points

The most serious mismatch, emphasized by Ding and Wright [3], concerns Lagrange multipliers. In the original problem (NLP), the KKT conditions impose $\lambda_i \geq 0$ for the inequality constraints $c_i(x) \geq 0$. In the SSV reformulation (SSV), the constraints $c_i(x) - s_i^2 = 0$ are equalities, so their multipliers $\mu_i \in \mathbb{R}$ are unconstrained in sign. A first-order stationary point of (SSV) with some $\mu_i < 0$ need not correspond to a KKT point of the original problem.

This is where naive first-order equivalence fails. The central message of the recent literature is that the correct comparison must be made at second order.

The three objections are closely linked. Loss of convexity creates additional stationary points; boundary degeneracy complicates their interpretation; and multiplier-sign ambiguity explains why first-order theory alone cannot resolve them. The next three sections revisit these issues through the second-order objects appropriate to each setting. In this order, the recent literature appears not as a collection of repairs but as a gradual identification of substitutes for the first-order picture that squaring destroys.

3 Benign Nonconvexity on the Sphere

3.1 Setup

The simplex provides the cleanest setting in the recent literature because the lifted search space remains a smooth manifold. Li, McKenzie, and Yin [8] study the simplex-constrained problem

$$\min_{x \in \Delta_n} f(x), \tag{C1}$$

where $\Delta_n = \{x \in \mathbb{R}^n : x \geq 0,\ \mathbf{1}^T x = 1\}$ is the standard probability simplex. The Hadamard parametrization $x = z \circ z$ with $\|z\|_2 = 1$ lifts (C1) to

$$\min_{z \in S^{n-1}} g(z) := f(z \circ z). \tag{NC1}$$

The map $\pi : S^{n-1} \to \Delta_n$, $\pi(z) = z \circ z$, is surjective, so the simplex is represented exactly as the image of the sphere under a coordinatewise squaring map. Interior points have 2^n preimages, corresponding to independent sign choices, whereas boundary points have fewer.

This setting isolates the first question behind the renewed interest in squared-variable parametrizations: can a lift introduce nonconvexity without creating spurious local minima? Because the sphere is a smooth, compact manifold, one can address that question before confronting the singular geometry that appears for more general feasible sets. The simplex case is therefore the cleanest place to see what the modern theory means by "benign" nonconvexity.

3.2 Critical-point correspondence

The conceptual core of [8] is that the Hadamard lift preserves the relevant distinction between optimality and negative curvature.

Theorem 3.1 (Li–McKenzie–Yin [8], informal synthesis)**.** *Under the map $\pi(z) = z \circ z$, KKT points of the simplex problem* (C1) *correspond to KKT points of the sphere problem* (NC1)*, and strict saddles correspond as well. Consequently, if f is convex on Δ_n, every first-order stationary point of* (NC1) *is either a strict saddle or a lift of a global minimizer of* (C1)*.*

The proof is constructive. The Riemannian gradient and Hessian of g on the sphere can be written explicitly, and their spectral behavior can then be compared directly with the KKT system of the original simplex problem. The result is therefore finer than a generic statement that there are no spurious local minima: it identifies which lifted stationary points encode optimality and which encode negative curvature.

The lift therefore changes the ambient geometry without changing the basic classification of critical points. The theorem has two conceptual layers. The KKT/strict-saddle correspondence is geometric and does not rely on convexity. Convexity enters only in the final step, where KKT points of the simplex problem coincide with global minimizers. This separation matters because it shows that lift-induced nonconvexity can be understood independently of global optimality.

3.3 Why this case matters

No spurious local minima in the convex case. The lifted problem is nonconvex, but every local minimizer maps to a global minimizer of the original simplex problem. In this setting, nonconvexity is introduced without creating a new class of harmful local minima.

A smooth manifold replaces a polyhedral feasible set. The reformulated problem lives on the sphere S^{n-1}, so one can use Riemannian optimization methods based on retractions and tangent-space operations rather than simplex projections. This is not merely a computational convenience; it is also what makes the geometric analysis clean.

Symmetry without ambiguity. The many-to-one nature of π over the simplex interior creates multiple lifted representatives of the same primal point, but these are symmetry copies rather than genuinely new primal solutions. This is one reason the additional stationary points remain interpretable.

Algorithmic consequences. Li, McKenzie, and Yin also study algorithms. They combine the parametrization with Riemannian methods that escape saddles and establish convergence to second-order points. They also prove local linear convergence for a projected Riemannian gradient scheme near nondegenerate solutions. Thus the theory is not only landscape-theoretic; it also explains why the lift can be useful in practice.

Example 3.2 (Running example)**.** Problem (Ex-H) on S^1 has stationary points at $y = (0, \pm 1)$ and $y = (\pm 1, 0)$. The former are global minima, mapping to $x^* = (0, 1)$; the latter are strict saddles, mapping to the nonoptimal vertex $(1, 0)$. Thus the nonconvexity is entirely benign in this simple example, exactly as predicted by Theorem 3.1.

4 Singular Geometry on Parametrized Polyhedra

4.1 Scope and setup

Tang and Toh [11] move from the simplex case to general polyhedra. They study (LCP) and its Hadamard parametrization (LCPH). The feasible set of the parametrized problem is

$$\mathcal{P}_{A,b} = \big\{ y \in \mathbb{R}^n : A(y \circ y) = b \big\}.$$

This set is the inverse image of b under the quadratic map $y \mapsto A(y \circ y)$, so it is a real algebraic variety defined by m polynomial equations of degree 2 in n variables.

The map is the same as before, but the geometry is not. The sphere is smooth; $\mathcal{P}_{A,b}$ is generally singular. If $\alpha := \{i : y_i^* = 0\} \neq \emptyset$, then the point y^* lies on a lower-dimensional stratum at which the usual manifold tangent space is no longer the right first-order object. One must instead work with a tangent cone, and for second-order analysis with second-order tangent sets. At this point, smooth-manifold intuition is no longer sufficient, and singular variational geometry becomes necessary.

Remark 4.1 (Scope limitation). The results of [11] apply to linearly constrained problems with nonnegativity bounds. They do not directly cover nonlinear inequality constraints of the form $g_i(x) \leq 0$.

4.2 Explicit tangent cones

A central contribution of [11] is the derivation of explicit formulas for the tangent cone $T_{\mathcal{P}}(y^*)$ at any point $y^* \in \mathcal{P}_{A,b}$. Writing

$$\alpha = \{i : y_i^* = 0\}, \qquad \beta = \{1, \ldots, n\} \setminus \alpha,$$

the cone can be described in terms of the submatrices A_α and A_β together with the support pattern of y^*. The resulting description is partly linear and partly one-sided: the components on the support β satisfy linear relations, whereas the zero coordinates indexed by α are governed by sign-type restrictions.

This explicitness makes the Tang–Toh approach distinctively algebraic. Because the feasible set is defined by polynomial equations, its local first-order geometry can be read from the support pattern of y^* and the combinatorial structure of A, rather than treated only through abstract variational analysis.

4.3 Second-order tangent sets

For singular problems, tangent cones alone are insufficient. Tang and Toh therefore derive explicit second-order tangent sets for $\mathcal{P}_{A,b}$. These objects encode how the variety bends along feasible first-order directions and play the role of the constraint-map Hessian on a smooth manifold.

This second-order information is essential near singular points. It is precisely what allows one to distinguish a genuine local minimizer from a point that is merely first-order feasible in a highly degenerate geometry.

4.4 Multiplier recovery

The theoretical payoff is a multiplier-recovery theory for the original polyhedral problem.

Theorem 4.2 (Tang–Toh [11], informal synthesis)**.** *Let y^* be feasible for* (LCPH)*, and set $x^* = y^* \circ y^*$. Explicit formulas for the tangent cone $T_{\mathcal{P}}(y^*)$ and second-order tangent sets $T^2_{\mathcal{P}}(y^*, h)$ permit one to recover a multiplier certificate for the original problem at x^*. In particular, every second-order stationary point of* (LCPH) *maps to a first-order stationary point of* (LCP)*. The construction does not require additional constraint qualifications at singular points.*

The significance is methodological: losing standard smooth-manifold structure changes the objects used for optimality certification but does not preclude certification. Singularity remains, but its structure becomes explicit enough to compute with. In the convex cases motivating much of the literature, the recovered first-order certificate is already enough to identify a global minimizer of the original polyhedral problem. More generally, tangent cones and second-order tangent sets replace the tangent-space and Hessian calculus available on smooth manifolds. This is the point at which the survey turns from benign nonconvexity to controlled singularity.

4.5 Stratification and hybrid algorithms

Tang and Toh further decompose $\mathcal{P}_{A,b}$ into finitely many smooth strata indexed by support patterns:

$$\mathcal{P}_{A,b} = \bigsqcup_{\alpha \subseteq \{1,\ldots,n\}} \mathcal{S}_\alpha, \qquad \mathcal{S}_\alpha = \big\{ y \in \mathcal{P}_{A,b} : y_i = 0 \Leftrightarrow i \in \alpha \big\}.$$

Each stratum $\mathcal{S}_\alpha$ is a smooth Riemannian manifold, so standard manifold methods apply locally. The main difficulty is then not optimization *within* a stratum but motion *across* strata when support changes occur.

Their proposed hybrid method reflects exactly this geometry: it performs Riemannian optimization while the iterate remains in a fixed stratum, detects impending support changes when a coordinate approaches zero, and switches to projected-gradient steps to traverse the resulting nonsmooth transition. The algorithm is not ad hoc: it follows from the explicit geometry of the lifted feasible set. Once the lift becomes singular, stratified second-order geometry provides a systematic alternative to case-by-case extensions of manifold calculus.

5 Second-Order Transfer

5.1 Scope

Ding and Wright [3] address the point at which the theory must become both more general and quantitative. Their paper studies direct square substitution for bound-constrained problems and squared-slack reformulations for general nonlinear programs. For survey purposes, the unifying theme is transfer: which second-order properties of the reformulated problem can be transferred back to the original inequality-constrained problem, and with what loss?

This is also the point at which exact geometric pictures are no longer enough. Once one asks about evaluation complexity or the output of a finite algorithm, the relevant object is approximate second-order optimality. The question is therefore not merely whether a second-order point transfers, but how the transferred residuals depend on the approximately active set and on the tolerance to which the reformulated problem has been solved.

5.2 Exact and approximate transfer

Ding and Wright compare the formulations at the correct level: second-order necessary conditions and their approximate variants, not first-order stationarity. It is useful to distinguish the exact DSS theory for bound constraints from the exact and approximate SSV theory for general nonlinear programs; a single slogan obscures the structure of the result.

Theorem 5.1 (Ding–Wright [3], informal synthesis). *The paper identifies three complementary transfer principles.*

(i) For bound-constrained problems under direct square substitution, exact second-order necessary points of the reformulation correspond to weak *second-order necessary points of the original problem; under strict complementarity, second-order sufficient conditions also correspond.*

(ii) For general nonlinear programs under squared-slack variables, exact second-order points of the reformulation correspond to weak *second-order points of the original problem. Thus the exact SSV theory already singles out weak second-order necessity as the natural target, rather than the full copositivity-based second-order condition.*

(iii) *For approximate squared-slack points, one selects an approximately active set using a threshold parameter ζ and transfers residuals componentwise—stationarity, feasibility, complementarity, primal–dual feasibility, and curvature. When ζ is held fixed and the reformulation is solved to accuracy ϵ, the transferred residuals are $O(\epsilon)$.*

Weak second-order conditions are the natural exact target. The first two items already show why second order is the correct level of comparison. Neither DSS nor SSV recovers the full second-order necessary condition of the original problem in complete generality; rather, the exact correspondences yield weak second-order necessity. This is not a technical nuisance. It records precisely how the squaring map replaces cone-based curvature tests in the original variables by curvature tests posed in the lifted variables.

The approximate SSV result depends on the threshold. The third item is best understood as an active-set-sensitive transfer theorem. One first fixes an approximately active set through the threshold ζ and then reconstructs multipliers with error terms that depend quantitatively on that choice. This is why a one-line summary such as "ϵ becomes $O(\sqrt{\epsilon})$" is only a rough mnemonic. The actual bounds are componentwise and depend on the regime in which ζ is chosen: if ζ is fixed, the transferred residuals are $O(\epsilon)$; if ζ shrinks with the tolerance, the more singular terms involving $\epsilon_1/\sqrt{\zeta}$ and $\epsilon_1/\zeta^{3/2}$ become visible.

Why second order is unavoidable. First-order stationarity is too weak because equality multipliers in the squared-slack problem need not have the sign required by the original KKT system. Second-order conditions detect the negative curvature associated with wrong-sign multipliers and thereby restore a meaningful correspondence.

A useful calculation makes this point transparent. With the sign convention above, write the squared-slack Lagrangian locally as

$$\mathcal{L}_{\mathrm{SSV}}(x,s,\mu) = f(x) - \sum_i \mu_i\big(c_i(x) - s_i^2\big).$$

At an active constraint, $s_i = 0$, the slack direction is invisible to the linearized constraint because $\partial(c_i(x) - s_i^2)/\partial s_i = 0$. Yet the second derivative of $\mathcal{L}_{\mathrm{SSV}}$ in the s_i direction is $2\mu_i$. Thus a multiplier with the wrong sign, $\mu_i < 0$, creates negative curvature along a direction that first-order analysis cannot exclude. This is the local mechanism behind the second-order transfer theorems.

Why the approximate theory matters. The theory does not merely assert that the two formulations remain "close." It identifies which residuals transfer and how their magnitudes depend on the approximately active set. This is the information needed to carry evaluation-complexity guarantees from equality-constrained algorithms back to the original inequality-constrained problem.

5.3 Algorithmic implications

Ding and Wright also examine the practical competitiveness of squared-variable formulations on bound-constrained quadratic programs, linear programs, and nonnegative matrix factorization. Their results do not show that squared-variable methods dominate specialized solvers; rather, they show that such methods can be computationally credible when paired with algorithms adapted to the lifted structure. For linear programming, the paper also draws suggestive parallels between the squared-variable viewpoint and primal–dual interior-point methods.

5.4 Constraint qualifications under SSV

The same paper also clarifies a point that is often misstated in informal discussions: LICQ is not automatically destroyed by squared slacks. At corresponding feasible points, LICQ for the original nonlinear program is equivalent to LICQ for the squared-slack reformulation [3]: once inactive constraints contribute nonsingular diagonal slack blocks, full row rank of the lifted Jacobian reduces to full row rank of the active constraint gradients. Because all constraints in SSV are equalities, the Mangasarian–Fromovitz constraint qualification (MFCQ) for the squared-slack problem is equivalent to LICQ there and therefore implies MFCQ for the original problem.

The practical subtlety lies elsewhere. Although LICQ may be preserved, the slack variables contribute no first-order information at active constraints, multiplier signs are no longer built into the reformulated KKT system, and the approximate theory must work with a thresholded active set. The main technical issues are therefore second-order and active-set-dependent rather than simple failures of classical regularity. One should not say that regularity is destroyed wholesale; it is redistributed across the active-set geometry of the lift.

6 Synthesis and Outlook

6.1 Three guiding questions

The recent literature can be organized less as a chronology than as three answers to three guiding questions. Table 1 summarizes the point.

Table 1: Recent work on squared-variable parametrizations.

Question	**Setting**	**Key work**	**Answer**
Does the lift create harmful critical points?	Convex simplex	Li–McKenzie–Yin	The extra stationary points are strict saddles
How should one certify optimality at singular lifted points?	Polyhedra	Tang–Toh	Tangent cones and second-order tangent sets recover a first-order certificate for the original problem
What quantitative guarantees survive under approximation?	General NLP	Ding–Wright	Approximate second-order residuals transfer with explicit active-set-dependent bounds

Together, these works show not that squaring is harmless but that its effects are highly structured. The singularity of $t \mapsto t^2$ appears exactly where boundary geometry matters, so the missing information is not random; it is displaced from first-order objects to second-order ones. First-order analysis alone is often too weak: multiplier signs may be wrong, slack columns may vanish exactly where complementarity matters, and singular support changes may be invisible. Once one passes to second-order objects—Riemannian Hessians, tangent cones and second-order tangent sets, or approximate second-order residuals—the correct optimality structure reemerges. The three theories surveyed above are therefore complementary rather than competing: each starts from the same singularity of $t \mapsto t^2$, but each identifies a different second-order object as the right viewpoint.

6.2 Connection to general parametrization theory

The squared-variable technique is a special case of a broader question: *when does a smooth reparametrization preserve the optimization landscape of a constrained problem?* Levin, Kileel, and Boumal [7] develop a general framework for studying how parametrizations affect critical points, local minima, and strict-saddle structure. In that sense, the Hadamard map is not an isolated trick but a concrete instance of a broader landscape-preservation theory for smooth, possibly nonconvex, lifts.

From this perspective, Tang and Toh's work is a refinement tailored to a singular polynomial parametrization. The map $y \mapsto y \circ y$ is smooth but not immersive at coordinates that vanish, so the general smooth-parametrization theory must be augmented by tools from algebraic geometry and variational analysis. This is exactly where tangent cones, second-order tangent sets, and stratification enter.

A related and better-known example is the Burer–Monteiro parametrization for semidefinite programming, which represents $X \succeq 0$ as $X = RR^T$. Ding and Wright [4] extend the squared-variable viewpoint to this matrix setting and obtain analogous second-order correspondences for nonlinear semidefinite programs.

6.3 The numerical price of squaring

The exact and approximate transfer theorems should not be confused with a promise of favorable conditioning. The map $y \mapsto y \circ y$ is singular whenever some coordinate vanishes, and that singularity has direct algorithmic consequences.

Gradient attenuation. For the lifted objective $F(y) = f(y \circ y)$, with $x = y \circ y$, the chain rule gives

$$\nabla F(y) = 2\ \mathrm{diag}(y)\, \nabla f(x).$$

Thus, whenever y_i is small, the corresponding gradient component in the lifted variables is attenuated, even if the original derivative $\partial f / \partial x_i$ is not. Near support changes or boundary solutions, first-order methods may therefore progress slowly in the coordinates that matter most.

Anisotropic curvature. The Hessian of the lifted objective is

$$\nabla^2 F(y) = 2\ \mathrm{diag}(\nabla f(x)) + 4\ \mathrm{diag}(y)\, \nabla^2 f(x)\ \mathrm{diag}(y).$$

The second term collapses in directions involving nearly vanishing coordinates, whereas the first term is controlled by first-order data in the original variables. Near the boundary, this mixture can create severe anisotropy and poor scaling, even when the original problem is well behaved.

These effects are the objective-side analogue of the boundary degeneracy discussed in Section 2.2: the same zero derivative of the squaring map that erases slack-column information in the constraint Jacobian also attenuates sensitivity in the lifted variables. This observation does not by itself characterize the conditioning of the full constrained KKT system, but it explains why boundary coordinates can become numerically difficult. The literature has begun to explain this phenomenon qualitatively, but a systematic condition-number theory for squared-variable parametrizations is still missing.

Algorithmically, these observations favor methods that use curvature information or identify active sets, including trust-region procedures, carefully designed second-order schemes, and the hybrid Riemannian/projected strategy of Tang and Toh [11].

6.4 Parametrization-induced nonconvexity

A broader lesson of the recent literature is that nonconvexity alone is not a decisive objection. What matters is the structure of the critical set: whether extra stationary points are strict saddles, benign symmetry copies, or genuinely spurious local minima. Much of modern optimization can be read through this lens, especially in settings where a smooth parametrization simplifies the feasible set at the cost of introducing redundancy.

The Hadamard lift offers a particularly clean case study. It replaces a polyhedral feasible set by a smooth manifold and introduces sign symmetries, thereby creating a nonconvex problem whose additional stationary points remain interpretable. In the convex simplex setting, the lifted landscape still preserves the decisive optimality information. For survey purposes, this makes the squared-variable literature a useful bridge between classical constrained optimization and the broader theory of benign nonconvex reformulation.

This perspective also matters in composite models, where simplex or nonnegativity constraints appear as substructures inside a larger nonconvex architecture. In such settings, squared-variable parametrizations offer a way to encode feasibility directly in the model parametrization rather than impose it through repeated projections or auxiliary constrained subproblems.

6.5 Practical takeaway

The results surveyed here suggest a practical distinction. DSS is most natural when nonnegativity is explicit and the remaining constraints have exploitable structure, as in simplex or polyhedral models. In that setting the lift changes the feasible geometry, but the new geometry can be analyzed directly. SSV is more natural when one wishes to convert a general inequality-constrained problem into an equality-constrained one, especially for algorithms or software interfaces that are built around equality constraints. Its main liability is not merely nonconvexity but the loss of multiplier-sign information at first order.

Consequently, squared-variable parametrizations should be chosen for structural reasons rather than for cosmetic reformulation. They are most appealing when the lifted problem admits efficient unconstrained, Riemannian, or equality-constrained steps, when strict-saddle behavior can be exploited, or when an approximate second-order point is the algorithmic target. They are less attractive when the solution is expected to lie on a highly degenerate boundary face and the available algorithm is purely first order, because the attenuation and anisotropy discussed in Section 6.3 may dominate the theoretical advantages.

6.6 Open directions

Recent developments have partly answered several questions that are sometimes presented as open. The following directions isolate the issues that remain unresolved.

1. **Singular nonlinear lifts.** Birtea, Caşu, and Comănescu develop an active-set stratification and explicit first- and second-order tests for smooth nonlinear inequalities on the regular part of a quadratic-slack lift [1]. A natural unresolved direction is a Tang–Toh-type theory at singular or degenerate points: explicit tangent cones and second-order tangent sets, multiplier recovery without regularity, and algorithms that remain valid when the lifted constraint gradients are linearly dependent.

2. **Sharp transfer bounds.** In the approximate theory of Ding and Wright, the transferred residuals depend on the active-set threshold ζ through terms such as $\epsilon_1/\sqrt{\zeta}$ and $\epsilon_1/\zeta^{3/2}$. To our knowledge, the available theory does not establish matching lower bounds or a general improvement in the dependence on ζ. Establishing whether these bounds are sharp, or improving them under stronger regularity or problem structure, remains an open quantitative question.

3. **Nonlinear-programming extensions.** Ding and Wright already give a detailed algebraic comparison between squared-slack sequential quadratic programming (SSV–SQP) and primal–dual interior-point steps for linear programming [3]. A corresponding analysis for nonlinear programming, together with global convergence, stability, and practical safeguards for an SSV–SQP method, remains to be developed.

4. **Rates on stratified varieties.** Tang and Toh's hybrid algorithm already chooses submanifolds adaptively and uses projected-gradient steps to move between them [11]. Their analysis leaves two questions open: convergence rates that reflect the local stratified geometry and the incorporation of second-order information into the algorithmic framework.

5. **Quantitative and nonsmooth extensions.** The KKT and strict-saddle correspondences of Li, McKenzie, and Yin do not require convexity; convexity is needed only to identify the resulting second-order KKT points with global minimizers [8]. Open directions include generic quantitative transfer from approximate lifted stationarity, weaker conditions for local linear convergence, and proximal, subgradient, or stochastic variants for nonsmooth and finite-sum objectives.

6. **Conditioning-aware formulations.** Recent Kurdyka–Łojasiewicz (KL) analyses sharpen the local variational and convergence theory

for Hadamard and square transformations [9, 10], but they do not remove the boundary singularity. Indeed, if a differentiable scalar map $\phi : \mathbb{R} \to \mathbb{R}_+$ attains zero at t_0, then $\phi'(t_0) = 0$. The meaningful design question is therefore how to mitigate, rather than eliminate, this degeneracy—for example, through redundant, nonsmooth, piecewise-smooth, or preconditioned formulations—while preserving rigorous stationarity and landscape correspondences.

7 Conclusion

Recent work rehabilitates squared-variable reformulations by classifying, rather than dismissing, their defects. Squaring introduces symmetry, nonconvexity, and boundary singularities, yet these features are now understood well enough to recover the relevant optimality structure in several important settings. On the simplex, Hadamard lifts preserve the decisive second-order and strict-saddle information and, under convexity, identify global solutions. For polyhedral constraints, tangent cones and second-order tangent sets permit analysis at singular lifted points. For general nonlinear programs, exact and approximate second-order conditions can be transferred between the original and reformulated problems under explicit assumptions.

The assessment must therefore remain conditional. These parametrizations are not mere tricks, but neither are they always preferable to the original formulations. Their value depends on the feasible geometry, the stationarity notion under study, and the intended algorithm. They are attractive when the lift yields tractable geometry or supports efficient unconstrained, Riemannian, or equality-constrained steps. They are less attractive when boundary degeneracy and numerical stiffness dominate.

More broadly, this literature illustrates a recurrent principle in modern optimization: introducing nonconvexity can simplify a constrained model without destroying its essential optimality structure, provided the lifted critical-point geometry remains sufficiently organized. Strict-saddle analysis, tangent geometry for singular varieties, and quantitative second-order transfer now make that principle precise for squared variables. The lasting lesson is therefore not that singular lifts are easy, but that they can be analyzed rigorously once the appropriate second-order and variational objects are brought into view.

References

[1] P. Birtea, I. Caşu and D. Comănescu, Optimization with inequality constraints by the embedded gradient vector field method, *arXiv preprint* arXiv:2606.20059, 2026. https://doi.org/10.48550/arXiv.2606.20059

[2] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, 2004.

[3] L. Ding and S.J. Wright, On squared-variable formulations, *SIAM Journal on Optimization*, **35** (2025), 2265–2293. https://doi.org/10.1137/23M1608343

[4] L. Ding and S.J. Wright, On squared-variable formulations for nonlinear semidefinite programming, *arXiv preprint* arXiv:2502.02099, 2025. https://doi.org/10.48550/arXiv.2502.02099

[5] E.H. Fukuda and M. Fukushima, A note on the squared slack variables technique for nonlinear optimization, *Journal of the Operations Research Society of Japan*, **60** (2017), 262–270. https://doi.org/10.15807/jorsj.60.262

[6] P.E. Gill, W. Murray and M.H. Wright, *Practical Optimization*, Academic Press, London, 1981.

[7] E. Levin, J. Kileel and N. Boumal, The effect of smooth parametrizations on nonconvex optimization landscapes, *Mathematical Programming*, **209** (2025), 63–111. https://doi.org/10.1007/s10107-024-02058-3

[8] Q. Li, D. McKenzie and W. Yin, From the simplex to the sphere: faster constrained optimization using the Hadamard parametrization, *Information and Inference: A Journal of the IMA*, **12** (2023), 1898–1937. https://doi.org/10.1093/imaiai/iaad017

[9] W. Ouyang, Kurdyka–Łojasiewicz exponent via square transformation, *arXiv preprint* arXiv:2506.10110, 2025. https://doi.org/10.48550/arXiv.2506.10110

[10] W. Ouyang, Y. Liu, T.K. Pong and H. Wang, Kurdyka–Łojasiewicz exponent via Hadamard parametrization, *SIAM Journal on Optimization*, **35** (2025), 62–91. https://doi.org/10.1137/24M1636186

[11] T. Tang and K.-C. Toh, Optimization over convex polyhedra via Hadamard parametrizations, *Mathematical Programming*, **214** (2025), 91–131. https://doi.org/10.1007/s10107-024-02172-2

Received: July 5, 2026; Published: July 20, 2026