Journal of Optimization Theory and Applications (2025) 204:33
https://doi.org/10.1007/s10957-024-02573-5

An Efficient Implementation of Interior-Point Methods for a Class of Nonsymmetric Cones

Yuwen Chen[1] · Paul Goulart[1]

Received: 14 July 2023 / Accepted: 13 November 2024

Abstract
We present an implementation of interior-point methods for generalized power cones, power mean cones and relative entropy cones, by exploiting underlying low-rank and sparsity properties of the Hessians of their logarithmically homogeneous self-concordant barrier functions. We prove that the augmented linear system in our interior-point method can be sparse and quasidefinite after adding a static regularization term, enabling the use of sparse LDL factorization for nonsymmetric cones. Numerical results show that our implementation can exploit sparsity in our test examples.

Keywords Nonsymmetric cone · Interior-point method · LDL factorization

1 Introduction

1.1 Literature Review

Convex optimization has been widely used in fields such as control engineering [3], signal processing [17] and finance [8]. It encompasses common problem classes such as linear programming (LP), quadratic programming (QP), second-order cone programming (SOCP) and semidefinite programming (SDP) problems, as well as other problems classes defined over more exotic cone constraints. Common modern approaches for general convex optimization include first-order operator-splitting methods [13, 25], and second-order interior-point methods (IPM) [22], where the latter is generally preferred when high accuracy solutions are required. Interior-point methods commonly employ ν-logarithmically homogeneous self-concordant barrier (LHSCB) functions, which apply penalties on convex inequality constraints defined

Communicated by Alper Yildirim.

✉ Yuwen Chen
yuwen.chen@eng.ox.ac.uk

Paul Goulart
paul.goulart@eng.ox.ac.uk

1 Department of Engineering Science, University of Oxford, Parks Road, OX1 3PJ, Oxford, UK

Published online: 22 January 2025

over cones. The parameter ν is related to the convergence speed of an IPM, where the total iteration number is known to be $\mathcal{O}(\sqrt{\nu}\log(1/\epsilon))$ in theory.

State-of-the-art conic interior-point solvers commonly employ an infeasible starting-point method where the initial iterate is chosen on the interior of any conic inequality constraints, with linear equalities not necessarily satisfied until the interior-point algorithm converges to optimality. Since general purpose optimization methods must allow for the possibility of infeasible problems being posed, IPMs require a mechanism that will work either by providing a solution (in the feasible case) or a certificate of infeasibility (in the infeasible case). This requirement motivated the development of the homogeneous self-dual embedding (HSDE) [34], which augments the original problem to a new one that is always feasible. The new augmented problem is always solvable, and from its solution one can recover either a solution or certificate of infeasibility for the original problem. The HSDE method is very popular and is widely used in both open-source and commercial solvers, e.g. SCS [25], CVXOPT [32], ECOS [12] and Mosek [19].

The great majority of research of conic IPMs focuses on problems with conic constraints that are homogeneous self-dual (also called self-scaled), e.g. the nonnegative, second-order and positive-semidefinite cones. Cones that do *not* possess the self-scaled property are called *nonsymmetric* and are topics of more recent interests for interior-point methods [23]. The exponential cone and the power cone are the two most commonly studied, and are supported in the commercial solver Mosek [19] and some open-source solvers like ECOS [12] and Clarabel [14]. It has also been shown that some existing conic optimization problems can be solved more efficiently by exploiting their special structure through the lens of nonsymmetric conic optimization, like sparse SDPs [1] and sum of squares (SOS) programs [26]. The Matlab-based DDS [16] and Julia-based Hypatia [6] solvers support a variety of nonsymmetric cone types that can solve many optimization problems more efficiently than their extended formulations based on three-dimensional exponential cones or power cones.

However, we cannot exploit Nesterov-Todd (NT) scaling points for nonsymmetric cones as already done in symmetric cones [21]. Instead a nonsymmetric strategy, where the scaling point is chosen to be the primal (dual) iterate in each iteration, is commonly used for nonsymmetric cones in solvers like ECOS and Hypatia. Recently, a symmetric primal-dual scaling algorithm motivated by [31] is proposed in [10] and implemented in Mosek and Clarabel [14]. Neighborhood check guarantees the update iterates close to the central path and some metrics require computing conjugate gradients of conic barrier functions, which generally do not have closed-form representations for nonsymmetric cones and efficient numerical methods have been proposed in [15].

The *power cone* is a powerful tool to model various cones such as the p-norm cone, the power mean cone and the generalized power cone [5]. The thesis of Chares [5] describes a 3-self-concordant barrier function for the standard three-dimensional power cone, and conjectured an $(n + 1)$-self-concordant barrier for the $(n + 1)$-dimensional power cone $\mathcal{K}_\alpha^{(n)} := \left\{(x, z) \in \mathbb{R}_+^n \times \mathbb{R} : \prod_{i=1}^n x_i^{\alpha_i} \geq |z|\right\}$, where $\alpha \in \mathbb{R}_\geq^n$ and $\sum_{i=1}^n \alpha_i = 1$. This was finally validated for a more general (n, m)-generalized power cone $\mathcal{K}_\alpha^{(n,m)} := \left\{(x, z) \in \mathbb{R}_+^n \times \mathbb{R}^m : \prod_{i=1}^n x_i^{\alpha_i} \geq \|z\|_2\right\}$ in [27]. The *relative* *entropy cone* can be regarded as the extension of three-dimensional exponential cone and is suitable to model problems in circuit design, statistical learning and power control in communication systems, etc. [4].

1.2 Organization

The paper is organized as follows: We introduce basic definitions of homogeneous self-dual embedding and interior-point methods in Sect. 2. We then propose a general structure that is suitable for sparse LDL factorization in Sect. 3. The proposed augmented sparse decomposition of Hessians for nonsymmetric cones are shown in Sect. 4. The implementation of the IPM is detailed in Sect. 5, and numerical results are given in Sect. 6. Finally, the conclusion is summarized in Sect. 7.

1.3 Notation

$\mathbb{R}^d$ is the d-dimensional space of reals and $\mathbb{R}^d_+$ ($\mathbb{R}^d_{++}$) denotes the nonnegative (positive) orthant. $\mathbb{S}^n$ is the set of symmetric matrices in $\mathbb{R}^{n\times n}$. Functions $f(\cdot)$ and $f^*(\cdot)$ consist of a convex conjugate pair. The kth-order derivative is denoted by $\nabla^k f(x)$. The set of interior points of $\mathcal{C}$ is denoted by $\operatorname{int}(\mathcal{C})$. The inner product of vectors x, y in the Euclidean space $\mathbb{R}^n$ is denoted by $\langle x, y\rangle$ and the Euclidean norm of x is denoted by $\|x\|$. The product set of $\mathcal{C}_1, \mathcal{C}_2$ is denoted as $\mathcal{C} := \mathcal{C}_1 \times \mathcal{C}_2$, e.g. $(x, y) \in \mathcal{C}$ means $x \in \mathcal{C}_1$, $y \in \mathcal{C}_2$. The index set $\{1, 2, \ldots, d\}$ is abbreviated as $[\![d]\!]$. An n-dimensional vector with value a for every entry is denoted by a^n. The indicator function $\delta_{i,j}$ is

$$\delta_{i,j} = \begin{cases} 1, & i = j, \\ 0, & \text{otherwise.} \end{cases}$$

$\operatorname{Diag}(x) \in \mathbb{R}^{n\times n}$ denotes a diagonal matrix where $x \in \mathbb{R}^n$ fills in the diagonal terms.

2 Problem Description

We will consider the following primal-dual pair of conic optimization problems throughout:

$$\begin{aligned} \min_{x,s} \quad & c^\top x \\ \text{s.t.} \quad & Gx = h, \\ & Ax + s = b, \; s \in \mathcal{K}, \end{aligned} \qquad (\mathcal{P})$$

$$\begin{aligned} \max_{y,z} \quad & -h^\top y - b^\top z \\ \text{s.t.} \quad & A^\top z + G^\top y + c = 0, \\ & z \in \mathcal{K}^*, \end{aligned} \qquad (\mathcal{D})$$

where $G \in \mathbb{R}^{p\times n}$, $A \in \mathbb{R}^{m\times n}$, $c \in \mathbb{R}^n$, $h \in \mathbb{R}^p$, $b \in \mathbb{R}^m$, $\mathcal{K}$ is a nonempty, closed convex cone and $\mathcal{K}^*$ is the dual cone of $\mathcal{K}$. We assume that $\mathcal{K}$ does not contain any straight lines, i.e. it is *pointed*. We call problem $\mathcal{P}$ the *primal* problem and $\mathcal{D}$ the *dual* problem.

The problem $\mathcal{P}$ and its dual $\mathcal{D}$ are quite general, and cover a wide range of standard problem types in numerical optimization. If $\mathcal{K}$ is the nonnegative orthant, i.e. $\mathcal{K} = \mathbb{R}_+$, then $\mathcal{P}$ represents a linear program. Likewise, if $\mathcal{K}$ is the second-order cone $\mathcal{K}_{\text{soc}}$ or the positive semidefinite cone $\mathcal{K}_{\text{sdp}}$, then $\mathcal{P}$ represents a second-order cone program or semidefinite program, respectively. In each of these cases, or where $\mathcal{K}$ is a composition of cones of these types, the cone $\mathcal{K}$ is *symmetric* and satisfies the self-dual property $\mathcal{K} = \mathcal{K}^*$.

In this paper we focus on problems in which the cone $\mathcal{K}$ is nonsymmetric and will consider in detail a family of nonsymmetric cones for which we can exploit special structure within an interior-point method. In the remainder of this section we sketch an interior-point method for such cones, and subsequently consider in detail the issue of sparsity exploitation for our cones of interest.

2.1 Logarithmically Homogeneous Self-Concordant Barrier (LHSCB) Functions

A function $f : \mathbb{R}^n \to \mathbb{R}$ is called $\nu-$LHSCB for a convex cone $\mathcal{K} \subset \mathbb{R}^n$, if it satisfies

$$\begin{aligned} &|\nabla^3 f(x)[r,r,r]| \leq 2\left(\nabla^2 f(x)[r,r]\right)^{3/2} && \forall x \in \operatorname{int}(\mathcal{K}),\, r \in \mathbb{R}^n, \\ &f(\lambda x) = f(x) - \nu \log(\lambda) && \forall x \in \operatorname{int}(\mathcal{K}),\, \lambda > 0, \end{aligned} \tag{2}$$

for a scalar $\nu \geq 1$. We first recall several properties of $\nu-$LHSCB functions and refer readers to [21] for more details.

The gradient $g : \mathbb{R}^n \to \mathbb{R}^n$ of a $\nu-$LHSCB f satisfies

$$\langle g(x), x\rangle = -\nu, \quad \forall x \in \operatorname{int}(\mathcal{K}). \tag{3}$$

The convex conjugate function $f^*(y) : \mathbb{R}^n \to \mathbb{R}$, defined as

$$f^*(y) := \sup_{x\in\operatorname{int}(\mathcal{K})} \{-\langle y, x\rangle - f(x)\}, \tag{4}$$

is also $\nu-$LHSCB in the interior of $\mathcal{K}^*$, and we call it the *conjugate barrier*. The gradient g^* of f^*, which we call the *conjugate gradient* of f, is the solution of

$$g^*(y) := -\arg\sup_{x\in\operatorname{int}(\mathcal{K})} \{-\langle y, x\rangle - f(x)\}. \tag{5}$$

The function f^* likewise has f and g as its conjugate and conjugate gradient, respectively. From (3)–(5) it can be verified that

$$f^*(y) = -\langle y, -g^*(y)\rangle - f(-g^*(y)) = -\nu - f(-g^*(y)),\ \forall y \in \operatorname{int}\left(\mathcal{K}^*\right). \tag{6}$$

For a $\nu-$LHSCB f, the gradient mapping plus the conjugate gradient mapping will eventually project a variable back to its original space, i.e.

$$-g^*(-g(x)) = x,\ \forall x \in \operatorname{int}(\mathcal{K}), \quad -g(-g^*(y)) = y,\ \forall y \in \operatorname{int}\left(\mathcal{K}^*\right). \tag{7}$$

2.2 Homogeneous Self-Dual IPM for Nonsymmetric Cones

We next introduce several basic building blocks for an IPM based on the HSDE for nonsymmetric cones.

In the standard HSDE model, the KKT condition for $\mathcal{P}$ and $\mathcal{D}$ are modified to

$$\begin{bmatrix} 0 \\ 0 \\ s \\ \kappa \end{bmatrix} = \begin{bmatrix} 0 & G^\top & A^\top & c \\ -G & 0 & 0 & h \\ -A & 0 & 0 & b \\ -c^\top & -h^\top & -b^\top & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ \tau \end{bmatrix} \tag{8}$$
$$(s, z, \kappa, \tau) \in \mathcal{F}, x \in \mathbb{R}^n, \ \ y \in \mathbb{R}^p,$$

where $\mathcal{F} := \mathcal{K} \times \mathcal{K}^* \times \mathbb{R}_+ \times \mathbb{R}_+$ and the cone $\mathcal{K}$ may represent the composition of multiple standard cone types. We refer to (8) as an HSDE model for $\mathcal{P}$–$\mathcal{D}$ since the problem of finding a feasible point for (8) is dual to itself[1], and any nonzero feasible point $(x^*, y^*, z^*, s^*, \tau^*, \kappa^*)$ of (8) can be used to extract either a solution to $\mathcal{P}$–$\mathcal{D}$ or a certificate of infeasibility [29]. The IPM amounts to a Newton-like method for finding a feasible point for (8).

The optimality condition in a convex problem includes a complementary slackness condition, which is achieved in the zero limit of the *central path* in an IPM. For symmetric cones, e.g. nonnegative, second-order and positive-semidefinite cones, we can exploit the self-scaled property [21] that ensures a unique (symmetric) Nesterov-Todd (NT) scaling point [21] is obtained from the combination of primal and dual variables. In the case of nonsymmetric cones, the nonsymmetric scaling is a popular choice [1, 23, 29] and is implemented in ECOS [28]. A primal-dual scaling with a 3rd-order correction is also effective and is implemented in Mosek [10]. In this paper, we will employ a nonsymmetric strategy with the scaling at dual iterates so that we can exploit the low-rank information or sparsity of Hessians of barrier functions directly.

Given some initial point $(x^0, y^0, s^0, z^0, \tau^0, \kappa^0)$ with $(s^0, z^0) \in \operatorname{int}(\mathcal{K}) \times \operatorname{int}(\mathcal{K}^*)$, the central path is parametrized by $\mu \in (0, 1]$ and the gradient $g^*(z)$ of the barrier function $f^*(z)$, satisfying

$$\begin{bmatrix} 0 \\ 0 \\ s \\ \kappa \end{bmatrix} = \begin{bmatrix} 0 & G^\top & A^\top & c \\ -G & 0 & 0 & h \\ -A & 0 & 0 & b \\ -c^\top & -h^\top & -b^\top & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ \tau \end{bmatrix} + \mu \begin{bmatrix} q_x \\ q_y \\ q_z \\ q_\tau \end{bmatrix}, \tag{9}$$
$$(s, z, \kappa, \tau) \in \mathcal{F},$$
$$s = -\mu g^*(z), \quad \tau\kappa = \mu,$$

[1] NB: this does *not* require that the cone $\mathcal{K}$ should be self-dual.

where

$$q_x = -G^\top y^0 - A^\top z^0 - c\tau^0, \qquad q_y = Gx^0 - h\tau^0,$$
$$q_\tau = \kappa + c^\top x^0 + h^\top y^0 + b^\top z^0, \qquad q_z = s^0 + Ax^0 - b\tau^0.$$

For an IPM with nonsymmetric cones, we require that every iterate should remain close to the central path. Standard metrics for this distance include the proximity measure in [23, 29] and one defined in terms of shadow iterates [10, 31].

2.3 Computation of the KKT System

In order to solve (8) we adopt the widely-used affine-centering (predictor-corrector) framework due to its excellent practical performance. We provide implementation details for this method in Sect. 5. Note that the step equations for both the predictor and the corrector steps can be written in the unified general form [32]

$$\begin{bmatrix} 0 & G^\top & A^\top & c \\ -G & 0 & 0 & h \\ -A & 0 & 0 & b \\ -c^\top & -h^\top & -b^\top & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \\ \Delta \tau \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \\ \Delta s \\ \Delta \kappa \end{bmatrix} = \begin{bmatrix} d_x \\ d_y \\ d_z \\ d_\tau \end{bmatrix}, \tag{10}$$
$$\mu H^*(z)\Delta z + \Delta s = -d_s, \;\; \tau \Delta\kappa + \kappa \Delta\tau = -d_\kappa,$$

where only the choice of right-hand side term $d_x, d_y, d_z, d_s, d_\tau, d_\kappa$ differs between the affine and the centering step directions in a given iterate. The Hessian of the barrier f^* is denoted as $H^*(z) \equiv \nabla^2 f^*(z) : \mathbb{R}^n \to \mathbb{S}^n$, which is positive definite for $\forall z \in \text{int}(\mathcal{K}^*)$.

We typically expect the data matrices G and A to be sparse, particularly for large problems, so we adopt the same approach as [14, 32] to preserve this sparsity while reducing the system to one that we can perform a symmetric linear solve. We first substitute the last two equations into the first one to obtain

$$\begin{bmatrix} 0 & G^\top & A^\top & c \\ G & 0 & 0 & -h \\ A & 0 & -\mu H^*(z) & -b \\ -c^\top & -h^\top & -b^\top & \frac{\kappa}{\tau} \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \\ \Delta \tau \end{bmatrix} = \begin{bmatrix} d_x \\ -d_y \\ d_s - d_z \\ d_\tau - d_\kappa/\tau \end{bmatrix}.$$

This can then be solved by solving a symmetric linear system twice with different right-hand sides, i.e.

$$K \begin{bmatrix} x_1 \\ y_1 \\ z_1 \end{bmatrix} = \begin{bmatrix} d_x \\ -d_y \\ d_s - d_z \end{bmatrix}, \qquad K \begin{bmatrix} x_2 \\ y_2 \\ z_2 \end{bmatrix} = \begin{bmatrix} -c \\ h \\ b \end{bmatrix}, \tag{11}$$

where K is defined as

$$K := \begin{bmatrix} 0 & G^\top & A^\top \\ G & 0 & 0 \\ A & 0 & -H_\mu \end{bmatrix} \tag{12}$$

with $H_\mu := \mu H^*(z)$ for the nonsymmetric scaling. Only the lower right-hand block H_μ, which may be dense in general, would be different if we had instead employed a primal-dual scaling strategy [10]. Finally, we obtain the solution of (10) by first solving for $\Delta\tau$ as

$$\Delta\tau = \frac{d_\tau - d_\kappa/\tau + c^\top x_1 + h^\top y_1 + b^\top z_1}{\kappa/\tau - c^\top x_2 - h^\top y_2 - b^\top z_2},$$

followed by

$$\begin{aligned} \Delta x &= x_1 + \Delta\tau x_2, & \Delta s &= -d_s - H_\mu \Delta z, \\ \Delta y &= y_1 + \Delta\tau y_2, & \Delta\kappa &= -\left(d_\kappa + \kappa\Delta\tau\right)/\tau. \\ \Delta z &= z_1 + \Delta\tau z_2, & & \end{aligned}$$

Observe that we only need to factorize the matrix K once at each iteration before performing the two linear solves in (11). We utilize sparse LDL factorization [11] with static ordering to factorize the matrix K. We add a small regularization term to the diagonal of K so that it becomes symmetric quasidefinite and the sparse LDL factorization always exists with diagonal D even after any symmetric permutation P [33], i.e. $K = PLDL^\top P^\top$.

3 Augmented Sparsity

We will consider a class of nonsymmetric cones for which the term H_μ that appears as the lower right-hand block in (12) is the sum of a sparse matrix plus a few low-rank dense terms, and satisfies a certain quasidefiniteness condition.

Definition 1 Suppose the matrix H is of the form

$$H := D + \sum_{i=1}^{n_1} u_i u_i^\top - \sum_{j=1}^{n_2} v_i v_i^\top, \tag{13}$$

where $u_i, v_i \in \mathbb{R}^n$, $D \in \mathbb{S}^{n\times n}$ is sparse and $n_1, n_2 \ll n$. We will say that H is *augmented-sparse* if $D - \sum_{j=1}^{n_2} v_i v_i^\top \succ 0$. The corresponding *augmented sparse matrix* is

$$H_{\text{aug}} := \begin{bmatrix} D & V & U \\ V^\top & I_{n_2} & \\ U^\top & & -I_{n_1} \end{bmatrix},$$

with $V = [v_1, \ldots, v_{n_2}]$, $U = [u_1, \ldots, u_{n_1}]$.

Lemma 1 *If H is augmented-sparse then H_{aug} is quasidefinite.*

Proof A sufficient condition is for the upper left 2×2 block of H to be positive definite [33]. This follows immediately since the Schur complement $D - VV^T$ is positive definite by assumption.

If H_μ in (12) is augmented-sparse, we can solve (11) via an equivalent linear system. For example, the last row of the left equation in (11), i.e. $Ax - H_\mu z_1 = d_s - d_z$ becomes

$$\begin{bmatrix} A \\ 0 \\ 0 \end{bmatrix} \Delta x - \underbrace{\begin{bmatrix} D & V & U \\ V^\top & I_{n_2} & \\ U^\top & & -I_{n_1} \end{bmatrix}}_{(H_\mu)_{\text{aug}}} \begin{bmatrix} \Delta z_1 \\ t_v \\ t_u \end{bmatrix} = \begin{bmatrix} d_s - d_z \\ 0 \\ 0 \end{bmatrix} \tag{14}$$

with additional variables $t_u \in \mathbb{R}^{n_1}$, $t_v \in \mathbb{R}^{n_2}$, which is both larger and sparser than the equivalent term in (11). The same applies to the last row of the right equation in (11). For (14), the number of nonzero entries in $(H_\mu)_{\text{aug}}$ is $[\text{nnz}(D)+n(n_1+n_2+2)]$, which is much smaller than the number of entries in a dense Hessian form if $n_1, n_2 \ll n$. We therefore expect that an LDL factorization of $(H_\mu)_{\text{aug}}$ will have significantly sparser factors than would be obtained by direct factorization of H_μ. Such a property has been exploited for SOCPs in ECOS under the NT (primal-dual) scaling strategy [12]. We will show how this approach can also be utilized in the nonsymmetric scaling strategy for nonsymmetric cones.

4 Sparse Hessian Decompositions for Nonsymmetric Cones

In this section we describe three nonsymmetric cones that can be shown to have *augmented-sparse* structure in the sense of Definition 1. We introduce fundamental definitions and barrier functions for these nonsymmetric cones in Sect. 4.1, and then detail the underlying augmented-sparse structures in Sect. 4.2.

4.1 Nonsymmetric Cones

We will show that the sparse structure can be exploited in three nonsymmetric cones of interest: generalized power cones, power mean cones and relative entropy cones. They are defined as follows:

Definition 2 [Generalized power cone] The *generalized power cone* is parametrized by $\alpha \in \mathbb{R}^{d_1}_{++}$ such that $\sum_{i\in[\![d_1]\!]} \alpha_i = 1$ and is defined as

$$\mathcal{C}_{\text{gpow}(\alpha,d_1,d_2)} = \left\{ (u, w) \in \mathbb{R}^{d_1}_{+} \times \mathbb{R}^{d_2} : \prod_{i\in[\![d_1]\!]} u_i^{\alpha_i} \geq \|w\| \right\}. \tag{15a}$$

Its dual cone is

$$\mathcal{C}^*_{\text{gpow}(\alpha,d_1,d_2)} = \left\{ (u, w) \in \mathbb{R}^{d_1}_+ \times \mathbb{R}^{d_2} : \prod_{i\in[\![d_1]\!]} \left(\frac{u_i}{\alpha_i}\right)^{\alpha_i} \geq \|w\| \right\}. \tag{15b}$$

Note that the dual generalized power cone is a linear transformation of the primal generalized power cone.

Definition 3 [Power mean cone] The *power mean cone* is parametrized by $\alpha \in \mathbb{R}^{d_1}_{++}$ such that $\sum_{i\in[\![d]\!]} \alpha_i = 1$ and is defined as

$$\mathcal{C}_{\text{powm}(\alpha,d)} = \left\{ (u, w) \in \mathbb{R}^d_+ \times \mathbb{R} : \prod_{i\in[\![d]\!]} u_i^{\alpha_i} \geq w \right\}. \tag{16a}$$

Its dual cone is

$$\mathcal{C}^*_{\text{powm}(\alpha,d)} = \left\{ (u, w) \in \mathbb{R}^d_+ \times \mathbb{R}_- : \prod_{i\in[\![d]\!]} \left(\frac{u_i}{\alpha_i}\right)^{\alpha_i} \geq -w \right\}. \tag{16b}$$

Note that the geometric mean cone [6] is a special case of the power mean cone where $\alpha_i = \frac{1}{d}, \forall i \in [\![d]\!]$.

Definition 4 [Relative entropy cone] The *relative entropy cone* is defined as

$$\mathcal{C}_{\text{rel}(d)} = \text{cl} \left\{ (u, v, w) \in \mathbb{R} \times \mathbb{R}^d_{++} \times \mathbb{R}^d_{++} : u \geq \sum_{i\in[\![d]\!]} w_i \log\left(\frac{w_i}{v_i}\right) \right\}. \tag{17}$$

Its dual cone is

$$\mathcal{C}^*_{\text{rel}}(d) = \text{cl} \left\{ (u, v, w) \in \mathbb{R}_{++} \times \mathbb{R}^d_{++} \times \mathbb{R}^d : w_i \geq u \left(\log\left(\frac{u}{v_i}\right) - 1\right), \forall i \in [\![d]\!] \right\}. \tag{18}$$

A generalized power cone reduces to a second-order cone when $d_1 = 1$ while it looks similar to a power mean cone when $d_2 = 1$. This motivates us to consider whether we can exploit sparsity of these nonsymmetric cones using an approach similar to the one that was described for second-order cones in [12]. The answer is *yes* if we choose the nonsymmetric scaling strategy for these cones in a IPM.

We therefore choose the nonsymmetric scaling strategy in our IPM throughout the paper. We will formulate our problems such that the cone treated as the "dual" cone $\mathcal{K}^*$ in our general primal-dual problem pair $(\mathcal{P})$–$(\mathcal{D})$ is actually one of the cones $\mathcal{C}_{\text{gpow}(\alpha,d_1,d_2)}$, $\mathcal{C}_{\text{powm}(\alpha,d)}$ or $\mathcal{C}_{\text{rel}(d)}$ from Definitions 2–4. We consequently require ν-LHSCB conjugate barrier functions for their support within an interior-point method, which have been defined already in [6, 16]:

Definition 5 The barrier functions for generalized power cones, power mean cones and relative entropy cones are defined as follows:

1. The function

$$f_{\text{gpow}}(u, w) = -\log\left(\prod_{i\in[\![d_1]\!]} u_i^{2\alpha_i} - \|w\|^2\right) - \sum_{i\in[\![d_1]\!]} (1-\alpha_i)\log(u_i) \tag{19}$$

is a (d_1+1)-LHSCB function of $\mathcal{C}_{\text{gpow}\,(\alpha,d_1,d_2)}$ where $\alpha, u \in \mathbb{R}^{d_1}_{++}$, $w \in \mathbb{R}^{d_2}$.

2. The function

$$f_{\text{powm}}(u, w) = -\log\left(\prod_{i\in[\![d]\!]} u_i^{\alpha_i} - w\right) - \sum_{i\in[\![d]\!]} \log(u_i) \tag{20}$$

is a $(d+1)$-LHSCB function of $\mathcal{C}_{\text{powm}(\alpha,d)}$ where $\alpha, u \in \mathbb{R}^{d}_{++}$, $w \in \mathbb{R}$.

3. The function

$$f_{\text{rel}}(u, v, w) = -\log\left(u - \sum_{i\in[\![d]\!]} w_i \log\left(\frac{w_i}{v_i}\right)\right) - \sum_{i\in[\![d]\!]} (\log(v_i) + \log(w_i)) \tag{21}$$

is a $(2d+1)$-LHSCB function of $\mathcal{C}_{\text{rel}(d)}$ where $v, w \in \mathbb{R}^{d}_{++}$.

4.2 Sparsity Exploitation

For the nonsymmetric scaling strategy at iteration k, the scaling matrix H_μ is set to $H_\mu^k = \mu^k H^*(z^k)$, where $\mu^k := \langle s^k, z^k\rangle/\nu > 0$ is the centering parameter and $H^*(z^k)$ is the Hessian of a barrier function at $z^k \in \mathcal{K}^*$ with degree ν. Hence, *we can exploit the augmented-sparse structure of the scaling matrix H_μ^k as long as $H^*(z^k)$ is augmented-sparse*. For the remainder of this section, we will show that the Hessians of barrier functions defined in Definition 5 are augmented-sparse or can be made augmented-sparse by adding a positive regularization to their diagonal, and thus so are the related scaling matrices H_μ^k.

4.2.1 Generalized Power Cone

We start by establishing the augmented-sparse property for the Hessian of generalized power cones:

Theorem 2 *The Hessian of the barrier function $f^*(z) = f_{gpow}(u, w)$ defined in (19) for the generalized power cone satisfies Definition 1 with $n_1 = 1, n_2 = 2$, i.e.*

$$H^*(z) = D + pp^\top - qq^\top - rr^\top,$$

where $z =: (u, w)$ *and* $D - qq^\top - rr^\top \succ 0$. *The parameters* D, p, q, r *are given by*

$$D = \left[\begin{array}{c|c} \underbrace{\begin{matrix} \ddots & & \\ & \frac{\tau_i \varphi}{\zeta u_i} + \frac{1-\alpha_i}{u_i^2} & \\ & & \ddots \end{matrix}}_{D_1} & \\ \hline & \underbrace{\frac{2}{\zeta} \cdot I_{d_2}}_{D_2} \end{array}\right], \quad p = \begin{bmatrix} p_0 \cdot \frac{\tau}{\zeta} \\ p_1 \cdot \frac{w}{\zeta} \end{bmatrix}, \tag{22a}$$

$$q = \begin{bmatrix} q_0 \cdot \frac{\tau}{\zeta} \\ 0 \end{bmatrix}, \; r = \begin{bmatrix} 0 \\ r_1 \cdot \frac{w}{\zeta} \end{bmatrix},$$

with

$$p_0 = \sqrt{\frac{\varphi(\varphi + \|w\|^2)}{2}}, \qquad p_1 = -2\sqrt{\frac{2\varphi}{\varphi + \|w\|^2}}, \tag{22b}$$

$$q_0 = \sqrt{\frac{\zeta\varphi}{2}}, \qquad r_1 = 2\sqrt{\frac{\zeta}{\varphi + \|w\|^2}},$$

where $\varphi = \prod_{i \in [\![d_1]\!]} u_i^{2\alpha_i}$, $\tau_i = \frac{2\alpha_i}{u_i}, \forall i \in [\![d_1]\!]$, *and* $\zeta = \prod_{i \in [\![d_1]\!]} u_i^{2\alpha_i} - \|w\|^2$.

Proof of Theorem 2 The gradient and Hessian for $f_{\mathrm{gpow}}(u, w)$ (19) are given by

$$\begin{aligned}
\nabla_{u_i} f &= -\frac{\tau_i \varphi}{\zeta} - \frac{1-\alpha_i}{u_i}, \quad \forall i \in [\![d_1]\!], \\
\nabla_{w_i} f &= \frac{2w_i}{\zeta}, \quad \forall i \in [\![d_2]\!], \\
\nabla^2_{u_i, u_j} f &= \frac{\tau_i \tau_j \varphi}{\zeta}\left(\frac{\varphi}{\zeta} - 1\right) + \delta_{i,j}\left(\frac{\tau_i \varphi}{\zeta u_i} + \frac{1-\alpha_i}{u_i^2}\right), \quad \forall i, j \in [\![d_1]\!], \\
\nabla^2_{u_i, w_j} f &= -\frac{2\tau_i \varphi w_j}{\zeta^2}, \quad \forall i \in [\![d_1]\!], \forall j \in [\![d_2]\!], \\
\nabla^2_{w_i, w_j} f &= \frac{4w_i w_j}{\zeta^2} + \delta_{i,j}\frac{2}{\zeta}, \quad \forall i, j \in [\![d_2]\!].
\end{aligned}$$

p_0, p_1, q_0, r_1 should satisfy the conditions

$$p_0^2 - q_0^2 = \varphi(\varphi - \zeta) = \varphi\|w\|^2, \quad p_0 p_1 = -2\varphi, \quad p_1^2 - r_1^2 = 4,$$

to match coefficients in the Hessian. Given the parameter choices in (22), we find $D - qq^\top - rr^\top$ is 2×2 block-diagonal and the condition $D - qq^\top - rr^\top \succ 0$

becomes

$$D_1 - q_0^2 \left(\frac{\tau}{\zeta}\right) \cdot \left(\frac{\tau}{\zeta}\right)^\top \succ 0 \text{ and } D_2 - r_1^2 \left(\frac{w}{\zeta}\right) \cdot \left(\frac{w}{\zeta}\right)^\top \succ 0,$$

which is equivalent to

$$\begin{aligned} 1 - q_0^2 \left(\frac{\tau}{\zeta}\right)^\top D_1^{-1} \left(\frac{\tau}{\zeta}\right) &= 1 - \sum_{i \in [\![d_1]\!]} \frac{q_0^2 \cdot \frac{\tau_i^2}{\zeta^2}}{\frac{\tau_i \varphi}{\zeta u_i} + \frac{1-\alpha_i}{u_i^2}} \\ &= 1 - \sum_{i \in [\![d_1]\!]} \frac{4\alpha_i^2 q_0^2}{\zeta(2\alpha_i \varphi + (1-\alpha_i)\zeta)} > 0, \end{aligned} \tag{23a}$$

and

$$1 - r_1^2 \left(\frac{w}{\zeta}\right)^\top D_2^{-1} \left(\frac{w}{\zeta}\right) = 1 - r_1^2 \frac{\|w\|^2}{\zeta^2} \cdot \frac{\zeta}{2} > 0, \tag{23b}$$

due to their Schur complements. If we set $q_0 = \sqrt{\frac{\zeta\varphi}{2}}, r_1 = 2\sqrt{\frac{\zeta}{\varphi + \|w\|^2}}$, (23a) is satisfied by

$$1 - \sum_{i \in [\![d_1]\!]} \frac{4\alpha_i^2 q_0^2}{\zeta(2\alpha_i \varphi + (1-\alpha_i)\zeta)} > 1 - \sum_{i \in [\![d_1]\!]} \frac{4\alpha_i^2 q_0^2}{2\zeta\alpha_i\varphi} = 1 - \frac{2q_0^2}{\zeta\varphi} = 0,$$

and (23b) is satisfied by

$$1 - r_1^2 \frac{\|w\|^2}{\zeta^2} \cdot \frac{\zeta}{2} = 1 - \frac{2\|w\|^2}{\varphi + \|w\|^2} = \frac{\zeta}{\varphi + \|w\|^2} > 0.$$

Hence, $D - qq^\top - rr^\top \succ 0$. □

Since the scaling matrix H_μ is augmented-sparse, we can employ an expanded sparse linear system

$$\begin{bmatrix} A \\ 0 \\ 0 \\ 0 \end{bmatrix} \Delta x - \mu \underbrace{\begin{bmatrix} D & q & r & p \\ q^\top & 1 & 0 & 0 \\ r^\top & 0 & 1 & 0 \\ p^\top & 0 & 0 & -1 \end{bmatrix}}_{(H_\mu)_{\text{aug}}} \begin{bmatrix} \Delta z_1 \\ t_q \\ t_r \\ t_p \end{bmatrix} = \begin{bmatrix} d_s - d_z \\ 0 \\ 0 \\ 0 \end{bmatrix}, \tag{24}$$

in our interior-point step equation (11), where $(H_\mu)_{\text{aug}}$ is quasidefinite and strongly factorizable [33]. Instead of (24), we solve the linear system

$$\begin{bmatrix} A \\ 0 \\ 0 \\ 0 \end{bmatrix} \Delta x - \begin{bmatrix} \mu D & \sqrt{\mu}q & \sqrt{\mu}r & \sqrt{\mu}p \\ \sqrt{\mu}q^\top & 1 & 0 & 0 \\ \sqrt{\mu}r^\top & 0 & 1 & 0 \\ \sqrt{\mu}p^\top & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} \Delta z_1 \\ t_q' \\ t_r' \\ t_p' \end{bmatrix} = \begin{bmatrix} d_s - d_z \\ 0 \\ 0 \\ 0 \end{bmatrix}, \tag{25}$$

which yields the same $\Delta x, \Delta z_1$ as (24). During direct factorization we find that the pivoting for (25) is more numerically stable than for (24) since the magnitude of diagonal terms of $(H_\mu)_{\text{aug}}$ increases from μ to 1 and the ratio of off-diagonal to diagonal terms also decreases, e.g. q to $\sqrt{\mu}q$.

In addition, the memory requirement for the rank-3 update of generalized power cones is equivalent to a rank-2 update due to the presence of zeros in q, r.

4.2.2 Power Mean Cone

We next establish the augmented-sparse property for the power mean cone:

Theorem 3 *The Hessian of the barrier function $f^*(z) = f_{powm}(u, w)$ defined in (20) for the power mean cone satisfies Definition 1 with $n_1 = 1, n_2 = 1$ given a positive regularization $\rho > 0$, i.e.*

$$H^*(z) = D + pp^\top - qq^\top,$$

where $z =: (u, w)$ and $D + \rho I - qq^\top \succ 0$. The parameters D, p, q are given by

$$D = \left[\begin{array}{ccc|c} \ddots & & & \\ & \frac{\tau_i \varphi}{u_i} + \frac{1}{u_i^2} & & \\ & & \ddots & \\ \hline & & & 0 \end{array}\right], \quad p = \begin{bmatrix} p_0 \cdot \tau \\ p_1 \cdot \frac{1}{\zeta} \end{bmatrix}, \quad q = \begin{bmatrix} q_0 \cdot \tau \\ 0 \end{bmatrix}, \tag{26a}$$

with

$$p_0 = \varphi, \quad p_1 = -1, \quad q_0 = \sqrt{\zeta\varphi}, \tag{26b}$$

where we define $\varphi = \prod_{i \in [\![d]\!]} u_i^{\alpha_i}$, $\zeta = \varphi - w$ and $\tau \in \mathbb{R}^d$ with $\tau_i = \frac{\alpha_i}{u_i \zeta}, \forall i \in [\![d]\!]$.

Proof of Theorem 3 The gradient and Hessian of (20) are

$$
\begin{aligned}
\nabla_w f &= \frac{1}{\zeta}, \\
\nabla_{u_i} f &= -\varphi \tau_i - \frac{1}{u_i}, \\
\nabla^2_{w,w} f &= \frac{1}{\zeta^2}, \\
\nabla^2_{w,u_i} f &= -\frac{\varphi \tau_i}{\zeta}, \\
\nabla^2_{u_i,u_j} f &= \varphi \tau_i \tau_j w + \delta(i,j) \left(\frac{\varphi \tau_i}{u_i} + \frac{1}{u_i^2} \right),
\end{aligned}
$$

According to the Schur complement lemma, the condition $D + \rho I - qq^\top \succ 0$ is equivalent to $\rho > 0$ and

$$
\begin{aligned}
&1 - q_0^2 \tau^\top \left[\mathrm{Diag} \left(\frac{\tau_i \varphi}{u_i} + \frac{1}{u_i^2} + \rho \right) \right]^{-1} \tau \succ 0 \\
=&1 - \sum_{i \in [\![d]\!]} \frac{\varphi \zeta \tau_i^2}{\frac{\tau_i \varphi}{u_i} + \frac{1}{u_i^2} + \rho} > 1 - \sum_{i \in [\![d]\!]} \frac{\alpha_i^2}{\alpha_i + \frac{\zeta}{\varphi}} > 0,
\end{aligned}
$$

which is valid since $(u, w) \in \mathrm{int}\left(\mathcal{C}_{\mathrm{powm}(\alpha,d)}\right)$ implies $\zeta, \varphi > 0$. Hence, we have $D + \rho I - qq^\top \succ 0$. □

The value $\rho > 0$ is referred to as a *static regularization* term and is commonly used in matrix factorization [24]. Note that the augmented-sparse structure of power mean cones is similar to the case of generalized power cones in Theorem 2, but with $r = 0$. We can therefore use an expanded sparse linear system as (24) with a numerically stable variant similar to (25) for solving equations (11).

4.2.3 Relative Entropy Cone

We can also exploit the augmented-sparse property in relative entropy cones after adding a static regularization:

Theorem 4 *The Hessian of the barrier function* (21) $f^*(z) = f_{\mathrm{rel}}(u, v, w)$ *for the relative entropy cone satisfies Definition 1 with* $n_1 = 1, n_2 = 0$ *after a positive regularization* $\rho > 0$*, i.e.*

$$
H^*(z) = D + pp^\top,
$$

where $z =: (u, v, w)$ *and* $D \succeq 0$, $D + \rho I \succ 0$. *The parameters* D, p *are given by*

$$D = \left[\begin{array}{c|c|c} 0 & 0 & 0 \\ \hline 0 & D_{vv} & D_{vw} \\ \hline 0 & D_{wv} & D_{ww} \end{array}\right] \succeq 0, \;\; p = \begin{bmatrix} \frac{1}{\zeta} \\ \sigma \\ \tau \end{bmatrix}, \tag{27}$$

with

$$\begin{aligned} D_{vv} &= \operatorname{Diag}\left(\left[\cdots, \frac{\sigma_i}{v_i} + \frac{1}{v_i^2}, \cdots\right]\right), \\ D_{ww} &= \operatorname{Diag}\left(\left[\cdots, \frac{1}{\zeta w_i} + \frac{1}{w_i^2}, \cdots\right]\right), \\ D_{vw} = D_{wv} &= \operatorname{Diag}\left(\left[\cdots, -\frac{1}{\zeta v_i}, \cdots\right]\right), \end{aligned}$$

where $\zeta = u - \sum_{i \in [\![d]\!]} w_i \log\left(\frac{w_i}{v_i}\right)$, $\tau_i = -\zeta^{-1}\left(\log\left(\frac{w_i}{v_i}\right) + 1\right)$, $\forall i \in [\![d]\!]$, $\sigma_i = \frac{w_i}{\zeta v_i}$, $\forall i \in [\![d]\!]$.

Proof The gradient of the barrier function (21) is

$$\begin{aligned} \nabla_u f &= -\frac{1}{\zeta}, \\ \nabla_{v_i} f &= -\sigma_i - \frac{1}{v_i}, \\ \nabla_{w_i} f &= -\frac{1}{w_i} - \tau_i, \end{aligned}$$

with the Hessian

$$\begin{aligned} \nabla^2_{u,u} f &= \frac{1}{\zeta^2}, \\ \nabla^2_{u,v_i} f &= \frac{\sigma_i}{\zeta}, \\ \nabla^2_{u,w_i} f &= \frac{\tau_i}{\zeta}, \\ \nabla^2_{v_i,v_j} f &= \sigma_i \sigma_j + \delta(i, j)\left(\frac{\sigma_i}{v_i} + \frac{1}{v_i^2}\right), \\ \nabla^2_{v_i,w_j} f &= \sigma_i \tau_j - \delta(i, j)\frac{1}{\zeta v_i}, \\ \nabla^2_{w_i,w_j} f &= \tau_i \tau_j + \delta(i, j)\left(\frac{1}{\zeta w_i} + \frac{1}{w_i^2}\right). \end{aligned}$$

Since $D_{v,v}, D_{w,w} \succ 0$, we have

$$\begin{bmatrix} D_{vv} & D_{vw} \\ D_{wv} & D_{ww} \end{bmatrix} \succ 0 \Leftrightarrow D_{v,v} - D_{v,w} D_{w,w}^{-1} D_{w,v} \succ 0,$$

due to the Schur complement, which is equivalent to

$$\frac{\sigma_i}{v_i} + \frac{1}{v_i^2} - \frac{\frac{1}{\zeta^2 v_i^2}}{\frac{1}{\zeta w_i} + \frac{1}{w_i^2}} > \frac{\sigma_i}{v_i} + \frac{1}{v_i^2} - \frac{w_i}{\zeta v_i^2} = \frac{1}{v_i^2} > 0,\ \forall i \in [\![d]\!].$$

Hence, we can obtain $D \succeq 0$ and then $D + \rho I \succ 0$ for $\rho > 0$.

5 An IPM for Nonsymmetric Conic Optimization

We are now in a position to summarize the implementation of our IPM algorithm.

5.1 Initialization of s^0, z^0

The initialization of symmetric cones and three-dimensional nonsymmetric cones, i.e. power and exponential cones, follows the Mosek setting as discussed in [10], where $x^0 = 0$, $\kappa^0 = \tau^0 = 1$ and s^0, z^0 are set to

$$s^0 = z^0 = -g^*(z^0), \tag{28}$$

which are on the central path with $\mu^0 = 1$. This strategy holds for $\mathcal{C}_{\mathrm{gpow}(\alpha,d_1,d_2)}$, and we initialize s^0, z^0 by

$$s^0 = z^0 = \left(\left(\sqrt{1+\alpha_i}\right)_{i\in[\![d_1]\!]}, 0^{d_2} \right).$$

For $\mathcal{C}_{\mathrm{powm}(\alpha,d)}$ and $\mathcal{C}_{\mathrm{rel}(d)}$, we can not easily find s^0, z^0 satisfying (28), so we initialize s^0, z^0 as in Hypatia [6], where s^0, z^0 satisfy $s^0 = -g^*(z^0)$ and then $\mu^0 = (\langle s^0, z^0\rangle + \tau^0\kappa^0)/(\nu+1) = 1$.

5.2 Proximity Measurement

A common approach in interior point methods is to keep iterates $(x^k, s^k, z^k, \tau^k, \kappa^k)$ in the neighborhood of the central path (9), since the solution of (9) is exactly an optimal solution of the KKT system (8) when μ tends to 0 from the positive side. The neighborhood is defined as

$$\mathcal{N}(\beta_1) := \|s + \mu g^*(z)\|_z^* = \left\langle s + \mu g^*(z), (H^*)^{-1}(z)\left(s + \mu g^*(z)\right)\right\rangle^{1/2} \le \beta_1 \mu \tag{29}$$

in Hypatia [6] where $\beta_1 \in [0, 1]$, and $\beta_1 = 0$ implies the iterate is on the central path. In Mosek's paper [10] with higher-order correction, the neighborhood is based on *shadow iterates* [31] with

$$\tilde{s} = -g^*(z), \quad \tilde{z} = -g(s), \quad \tilde{\mu} = \langle \tilde{s}, \tilde{z} \rangle / \nu,$$

for $(s, z) \in \mathcal{K} \times \mathcal{K}^*$, where $\mu\tilde{\mu} \geq 1$ with equality only on the central path [22], and the neighborhood is defined as

$$\mathcal{N}(\beta_2) := \left\{ (s, z) \in \mathcal{K} \times \mathcal{K}^* \mid \beta_2 \mu \tilde{\mu} \leq 1 \right\}, \tag{30}$$

while $\beta_2 \in (0, 1]$ and $\beta_2 = 1$ implies the iterate is on the central path. We use the neighborhood $\mathcal{N}(\beta_2)$ defined in (30) for generalized power cones and power mean cones, and use (29) for relative entropy cones [6] since the conjugate gradient of relative entropy cones can not be obtained efficiently.

5.3 Higher-Order Correction

The higher-order correction is an effective technique accelerating the convergence of IPMs in practice, and we use the third-order correction initially proposed in [10], which becomes

$$\eta(\Delta s_a, \Delta z_a) := -\frac{1}{2} \nabla^3 f^*(z) \left[\Delta z_a, \left(\nabla^2 f^*(z) \right)^{-1} \Delta s_a \right], \tag{31}$$

in our implementation, where Δs_a, Δz_a are affine directions detailed later in Sect. 5.4. Although a third-order correction was proposed for use in a symmetric scaling method [10] and the convergence proof does not hold for an IPM with nonsymmetric scaling, we find it still reduces computational time of the IPM with nonsymmetric scaling in numerical experiments. The computation of third-order derivatives and the inverse Hessians are detailed in Appendix A.

5.4 Algorithmic Sketch

Suppose we define the linear residual as

$$\mathcal{R}(x, y, z, s, \tau, \kappa) := \begin{bmatrix} 0 & G^\top & A^\top & c \\ -G & 0 & 0 & h \\ -A & 0 & 0 & b \\ -c^\top & -h^\top & -b^\top & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ \tau \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \\ s \\ \kappa \end{bmatrix}.$$

Our IPM proceeds as follows at each iteration k:

1. *Update residuals, gap and check termination or infeasibility:* We compute the residual $\mathcal{R}(x^k, y^k, z^k, s^k, \tau^k, \kappa^k)$ and $\mu^k = \frac{{s^k}^\top z^k + \kappa^k \tau^k}{\nu+1}$. The computation of duality gap, the termination check and the infeasibility detection follow ECOS [12].

2. *Affine step:* The affine direction (predictor) is the solution of

$$
\begin{aligned}
&\mathcal{R}(\Delta x_a, \Delta y_a, \Delta z_a, \Delta s_a, \Delta \tau_a, \Delta \kappa_a) = -\mathcal{R}(x^k, y^k, z^k, s^k, \tau^k, \kappa^k), \\
&\mu^k H^*(z^k)\Delta z_a + \Delta s_a = -s^k, \quad \tau^k \Delta \kappa_a + \kappa^k \Delta \tau_a = -\tau^k \kappa^k,
\end{aligned} \tag{32}
$$

   which tries to remove residuals of the linearized model at iteration k. We can compute the maximal step size α_a such that $(s^k + \alpha_a \Delta s_a, z^k + \alpha_a \Delta z_a, \tau^k + \alpha_a \Delta \tau_a, \kappa^k + \alpha_a \Delta \kappa_a)$ resides in $\mathcal{F}$.
3. *Update the sparse decomposition of $H^*(z)$*: We update the augmented sparse decomposition of the Hessian $H^*(z)$ for nonsymmetric cones as discussed in Sect. 3 and Sect. 4.
4. *Combined step:* The weight of centrality σ is set to $(1-\alpha_a)^3$ empirically and then used for the computation of the centering direction (corrector) via

$$
\begin{aligned}
&\mathcal{R}(\Delta x_c, \Delta y_c, \Delta z_c, \Delta s_c, \Delta \tau_c, \Delta \kappa_c) = -(1-\sigma)\mathcal{R}(x^k, y^k, z^k, s^k, \tau^k, \kappa^k), \\
&\mu^k H^*(z^k)\Delta z_c + \Delta s_c = -s^k - \sigma \mu^k g^*(z^k) - \eta(\Delta s_a, \Delta z_a), \\
&\tau^k \Delta \kappa_c + \kappa^k \Delta \tau_c = -\tau^k \kappa^k + \sigma \mu^k - \Delta \tau_a \Delta \kappa_a,
\end{aligned} \tag{33}
$$

   where $\Delta \tau_a \Delta \kappa_a$ and the function $\eta(\Delta s_a, \Delta z_a)$ are higher-order corrections given information from the affine directions. We use the Mehrotra's correction [18] for symmetric cones and the third-order correction (31) for nonsymmetric cones. Likewise, we compute the largest step size α_c ensuring $(s^k + \alpha_c \Delta s_c, z^k + \alpha_c \Delta z_c, \tau^k + \alpha_c \Delta \tau_c, \kappa^k + \alpha_c \Delta \kappa_c)$ stays inside $\mathcal{F}$ and a neighborhood $\mathcal{N}(\beta)$ of the central path.
5. *Update iterates:* At the end of each iteration k, we obtain the new iterate $(x^{k+1}, y^{k+1}, s^{k+1}, z^{k+1}, \kappa^{k+1}, \tau^{k+1})$ by

$$
\begin{aligned}
(x^{k+1}, y^{k+1}, s^{k+1}, z^{k+1}, \kappa^{k+1}, \tau^{k+1}) &:= (x^k, y^k, s^k, z^k, \kappa^k, \tau^k) \\
&+ \alpha_c(\Delta x_c, \Delta y_c, \Delta s_c, \Delta z_c, \Delta \kappa_c, \Delta \tau_c).
\end{aligned}
$$

6 Experiments

In our experiments, we first choose two examples to show the effectiveness of our sparse implementation for generalized power cones and power mean cones; the maximum likelihood estimation of a convex distribution [30] and the maximum volume hypercube problems [7]. We test these examples with 4 different solver configurations:

1. `GenPow` is the proposed sparse implementation of generalized power cones;
2. `PowMean` is the proposed sparse implementation of power mean cones;
3. `Clarabel-Pow` uses Clarabel but transforms the generalized power cone to a product of three-dimensional power cones as formulated in [5, 19];
4. `Hypatia` uses the Hypatia solver, which supports the generalized power cone and the power mean cone constraints directly;
5. `Mosek` denotes the use of Mosek with transforming the generalized power cone to a product of three-dimensional power cones.

In order to compare performance of our method against solvers that do not directly support generalized power cone constraints, we apply a standard transformation to reformulate the constraint for those solvers into a collection of three-dimensional power cone constraints. Following [19], we can rewrite a generalized power cone $(x, t) \in \mathcal{C}_{\text{gpow}(\alpha,n,1)}$ constraint as a collection of constraints over three-dimensional power cones

$$\begin{aligned} & x_1^{1-p_2} x_2^{p_2} \geq |z_3|\,, \\ & z_3^{1-p_3} x_3^{p_3} \geq |z_4|\,, \\ & \cdots \\ & z_{n-1}^{1-p_{n-1}} x_{n-1}^{p_{n-1}} \geq |z_n|\,, \\ & z_n^{1-p_n} x_n^{p_n} \geq |t|, \end{aligned} \tag{34}$$

where $p_i = \alpha_i / (\sum_{j=1}^{i} \alpha_j)$. This is the form used by `Clarabel-Pow` and `Mosek`. For relative entropy cones, we compare our sparse implementation called `Entropy` to the counterpart supported in the Hypatia solver on the signomial optimization problem [20] and the Wasserstein barycenter problem [9]. The convergence tolerance is set to $\epsilon = 10^{-7}$ for all tests.[2]

6.1 Maximum Likelihood Estimator of a Density Function

Given an ordered sample $y_1 < y_2 < \ldots < y_n$ of n outcomes of an unknown distribution with a density function $g : \mathbb{R}_+ \to \mathbb{R}_+$, we can form an estimate of g using a piecewise linear function estimator $\hat{g} : [y_1, y_n] \to \mathbb{R}_+$ with break points at $(y_i, x_i)\,, i = 1, \ldots, n$, where the variables $x_i > 0$ are estimators for $g\,(y_i)$ [30]. The slope of the i-th linear segment of $\hat{g}$ is

$$\frac{x_{i+1} - x_i}{y_{i+1} - y_i}.$$

We can enforce the gradient change to be monotone, such as the non-decreasing slope constraints

$$\frac{x_{i+1} - x_i}{y_{i+1} - y_i} \leq \frac{x_{i+2} - x_{i+1}}{y_{i+2} - y_{i+1}}, i = 1, \ldots, n-2.$$

The discrete density function $\hat{g}$ must sum to 1, i.e.

$$\sum_{i=1}^{n-1} (y_{i+1} - y_i) \left(\frac{x_{i+1} + x_i}{2} \right) = 1.$$

[2] https://github.com/yuwenchen95/Clarabel-genpowcone.jl.git

Table 1 Likelihood estimation with the generalized power cone

	GenPow		Hypatia (GenPow)		Clarabel-Pow		Mosek	
n	Time	Iter	Time	Iter	Time	Iter	Time	Iter
465	0.042	41	1.037	57	0.103	45	0.078	24
910	0.088	49	8.175	137	0.217	52	0.297	44
1370	0.223	87	19.144	138	0.390	61	0.625	63

Table 2 Likelihood estimation with the power mean cone

	PowMean		Hypatia (PowMean)	
n	Time	Iter	Time	Iter
465	0.083	55	0.905	45
910	0.169	51	5.304	84
1370	0.335	86	15.274	111

The maximum likelihood estimation can then be formulated as a conic optimization problem

$$\begin{aligned} \max_{x,t} \quad & t \\ \text{s.t.} \quad & \frac{x_{i+1}-x_i}{y_{i+1}-y_i} - \frac{x_{i+2}-x_{i+1}}{y_{i+2}-y_{i+1}} \leq 0, \quad i = 1,\ldots,n-2, \\ & \sum_{i=1}^{n-1} (y_{i+1}-y_i)\left(\frac{x_{i+1}+x_i}{2}\right) = 1, \\ & (x,t) \in \mathcal{K},\ x \geq 0, \end{aligned} \tag{35}$$

where $\mathcal{K}$ can be either a generalized power cone or a power mean cone. We test implementation of `GenPow`, `PowMean` and compare them with the counterparts implemented in `Hypatia`, `Clarabel-Pow` and `Mosek`, which are tested by replacing the constraint $(x, t) \in \mathcal{K}$ with the equivalent reformulation shown in (34).

Computational results including the total computational time and the number of iterations for different dimensions n are shown in Tables 1 and 2.

Table 1 shows that our sparse LDL implementation for generalized power cones performs better than other methods on problems up to moderate size. Although the total iteration number of `GenPow` is larger than the implementation of `Clarabel-Pow` or `Mosek`, it takes less time overall for convergence due to improved linear solve efficiency enabled by the augmented sparse decomposition we proposed in Sect. 4.2.1.

Table 2 summarizes results for the same problem, but choosing $\mathcal{K}$ to be a power mean cone in (35). Since `Mosek` doesn't support power mean cones directly, we compare our sparse implementation only with Hypatia. We vary the dimension n as for the test of generalized power cones. Table 2 shows that our sparse implementation for power mean cones is significantly faster than Hypatia which is based on solving

Table 3 Volume maximization problem with the generalized power cone

	GenPow		Hypatia (GenPow)		Clarabel-Pow		Mosek	
n	Time(s)	Iter	Time(s)	Iter	Time(s)	Iter	Time(s)	Iter
100	0.014	20	0.075	13	0.017	20	0.016	10
500	0.079	49	0.722	18	0.062	24	0.046	12
2500	0.648	80	31.243	21	0.431	26	0.297	14

Table 4 Volume maximization problem with the power mean cone

	PowMean		Hypatia (PowMean)	
n	Time(s)	Iter	Time(s)	Iter
100	0.019	25	0.091	17
500	0.087	49	1.228	16
2500	1.865	63	33.297	21

the reduced normal system where the sparsity is usually lost, both in total time and the averaged time per iteration. Although the results of our power mean cones are slower than our sparse implementation of generalized power cones, it still performs better than other solvers on the discrete maximum likelihood problem.

6.2 Maximum Volume of a Hypercube

The problem of finding the maximum volume of a hypercube [2] that fits inside the intersection of 1-norm and ∞-norm balls is given as follows,

$$\begin{aligned} &\max \quad t \\ &\text{s.t.} \quad (x, t) \in \mathcal{K}, \alpha = \left[\frac{1}{n}, \ldots, \frac{1}{n}\right], \\ &\qquad Ax \in \mathcal{B}_1(\gamma_1),\ \ Ax \in \mathcal{B}_\infty(\gamma_2), \end{aligned} \tag{36}$$

where $\mathcal{K}$ is either a generalized power cone or a power mean cone, $\mathcal{B}_1(\gamma_1) := \{x \mid \|x\|_1 \leq \gamma_1\}$, $\mathcal{B}_\infty(\gamma_2) := \{x \mid \|x\|_\infty \leq \gamma_2\}$ and γ_1, γ_2 are given constants. Both norm constraints are easily converted into a collection of linear inequalities. The matrix A is set to the identity matrix and the solver settings are kept the same as in the previous example, along with the same transformation of the generalized power cone to a product of three-dimensional power cones as in (34).

We vary the dimension n from 100 to 2500 and summarize the computational results in Tables 3 and 4. Although `GenPow` and `PowMean` take more time than `Clarabel-Pow` and `Mosek` in total time, the averaged time per iteration in our sparse implementation is smaller. This demonstrates the efficacy of our sparse augmentation approach compared to `Hypatia` that doesn't exploit sparsity in the matrix factorization. Since both our IPM and `Hypatia` need to solve the same linear system

Table 5 Test results for the signomial minimization problem

m	Entropy		Hypatia (Entropy)	
	Time(s)	Iter	Time(s)	Iter
10	0.07	62	0.89	60
20	0.19	85	3.06	45
30	0.46	114	12.30	36
40	0.71	100	61.68	57
50	1.22	105	152.56	50

Table 6 Test results for the Wasserstein barycenter problem

(m,n)	Entropy		Hypatia (Entropy)	
	Time(s)	Iter	Time(s)	Iter
(10,5)	0.07	17	8.96	25
(12,6)	0.63	23	44.51	35
(14,7)	0.89	27	226.74	43
(16,8)	1.07	34	1186.69	51
(18,9)	1.81	36	-	-

at each iteration and the sparse implementation of the Hessian matrix is isolated from the remaining parts of an IPM, it could be exploited in Hypatia for acceleration.

6.3 Signomial Minimization Problem

We take the signomial minimization problem from Hypatia's benchmark test set, which originated from [20] and formulates the convex SAGE relaxation problem with relative entropy cones as:

$$\begin{aligned} \max_{\gamma\in\mathbb{R},\mu\in\mathbb{R}^q_+} \quad & \gamma \\ \text{s.t.} \quad & c-\gamma\mathbf{1}^m-\sum_{p\in[\![q]\!]}\mu_p g_p = t, \\ & t=\sum_{k\in[\![m]\!]} X_{:,k}, \\ & (A_{[\![m]\!]\backslash\{k\},j}-A_{k,j}\mathbf{1}^{m-1})^\top Y_{:,k}=0, \forall k\in[\![m]\!], \forall j\in[\![n]\!], \\ & \begin{bmatrix} X_{k,k}+\sum_{j\in[\![m-1]\!]} Y_{j,k};\, X_{[\![m]\!]\backslash\{k\},k};\, Y_{:,k} \end{bmatrix} \in \mathcal{C}_{\mathrm{rel}(m-1)}, \forall k\in[\![m]\!], \end{aligned} \tag{37}$$

where $t\in\mathbb{R}^m$, $X\in\mathbb{R}^{m\times m}$, $Y\in\mathbb{R}^{(m-1)\times m}$ with parameters $c, g_p\in\mathbb{R}^m$, $p\in[\![q]\!]$ and $A\in\mathbb{R}^{m\times n}$. $[\![m]\!]\backslash\{k\}$ denotes the set difference between $[\![m]\!]$ and $\{k\}$.

We vary m from 10 to 50, which is the number of relative entropy cones. Table 5 shows that our IPM implementation can reduce the computational time by one to two orders of magnitude relative to the Hypatia solver.

6.4 Wasserstein Barycenter with Entropy Regularizations

Finally, we compare our implementation of relative entropy cones with Hypatia on the Wasserstein barycenter problem with entropy regularization, which is equivalent to a discrete optimal transport problem with the entropy regularization [9]:

$$
\begin{aligned}
\min_{p} \quad & \langle M, p\rangle + \frac{1}{\lambda}\sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij}\log\left(p_{ij}\right) \\
\text{st.} \quad & \sum_{j=1}^{m} p_{ij} = \mu_i \ \forall i \in [\![n]\!], \ \sum_{i=1}^{n} p_{ij} = \nu_j \ \forall j \in [\![m]\!], \\
& p_{ij} \geq 0 \quad \forall i \in [\![n]\!], \forall j \in [\![m]\!],
\end{aligned}
\tag{38}
$$

where $p \in \mathbb{R}^{n\times m}$ is the transport and $M \in \mathbb{R}^{n\times m}$ is a coefficient matrix. We vary the dimension of m, n and limit computation times to one hour in all tests. Table 6 shows that the time for Hypatia's implementation of relative entropy cones increases rapidly for problem (38), while our method finishes the computation within seconds. This suggests that the proposed sparsity exploitation for relative entropy cones can also be utilized within the Hypatia solver similar to generalized power cones and power mean cones.

7 Conclusion

We have presented an efficient sparse decomposition of generalized power cones, power mean cones and relative entropy cones within interior-point methods. We have detailed how to exploit the low-rank property inside Hessian matrices of these barrier functions and proposed the augmented decomposition that ensures the quasidefiniteness of the augmented linear system and makes it factorizable under the LDL factorization with static pivoting. The experimental results demonstrate that our sparse decomposition for these cones performs much better than the Hypatia solver and open the way to exploit the sparsity of nonsymmetric cones when they are supported directly within a solver.

Although the factorization time can be reduced dramatically, we observe that the number of iterations for convergence of high-dimensional nonsymmetric cones is higher compared to the equivalent problem solved by Mosek due to the smaller step size in each iteration, which is also found in tests of Hypatia. Future work would be designing more efficient and provable higher-order corrections, improving the numerical stability of IPMs and exploiting the low-rank property in the primal-dual scaling IPM for high-dimensional nonsymmetric cones.

Appendix A: Third-Order Derivatives and the Inverse of Hessian

We summarize the computation of third-order derivatives and the inverse Hessians for different cones following the derivation from Hypatia cones reference [6]:

A.1 Generalized Power Cone

Note that $\varphi = \prod_{i\in[\![d_1]\!]} u_i^{2\alpha_i}$, $\tau_i = \frac{2\alpha_i}{u_i}$, $\forall i \in [\![d_1]\!]$, and $\zeta = \prod_{i\in[\![d_1]\!]} u_i^{2\alpha_i} - \|w\|^2$. The third-order derivative is

$$\nabla^3_{u_i,u_j,u_k} f_{\text{gpow}} = \tau_i\tau_j\tau_k\frac{\varphi}{\zeta}\left(1-\frac{\varphi}{\zeta}\right)\left(\frac{2\varphi}{\zeta}-1\right) + \begin{cases} 2\frac{\tau_i\tau_j}{u_i}\frac{\varphi}{\zeta}\left(1-\frac{\varphi}{\zeta}\right), & i=j=k \\ \frac{\tau_i\tau_j}{u_i}\frac{\varphi}{\zeta}\left(1-\frac{\varphi}{\zeta}\right), & i=k\neq j \\ \frac{\tau_i\tau_j}{u_j}\frac{\varphi}{\zeta}\left(1-\frac{\varphi}{\zeta}\right), & i\neq j=k \\ 0, & \text{otherwise} \end{cases}$$

$$+\begin{cases} \frac{\tau_i^2\varphi}{u_i\zeta}\left(1-\frac{\varphi}{\zeta}\right) - (1-\alpha_i)\frac{2}{u_i^3} - \frac{2\varphi\tau_i}{\zeta u_i^2}, & i=j=k \\ \frac{\tau_i\tau_k\varphi}{u_i\zeta}\left(1-\frac{\varphi}{\zeta}\right), & i=j\neq k \\ 0, & \text{otherwise} \end{cases},$$

$$\nabla^3_{u_i,u_j,w_k} f_{\text{gpow}} = \frac{2\tau_i\tau_j\varphi w_k}{\zeta^2}\left(\frac{2\varphi}{\zeta}-1\right) + \delta(i,j)\frac{2\tau_i\varphi w_k}{\zeta^2 u_i},$$

$$\nabla^3_{u_i,w_j,w_k} f_{\text{gpow}} = \frac{-8w_jw_k\tau_i\varphi}{\zeta^3} - \delta(j,k)\frac{2}{\zeta^2}\tau_i\varphi,$$

$$\nabla^3_{w_i,w_j,w_k} f_{\text{gpow}} = \frac{16w_iw_jw_k}{\zeta^3} + \delta(i,j)\frac{4w_k}{\zeta^2} + \begin{cases} \frac{8w_i}{\zeta^2}, & i=j=k \\ \frac{4w_j}{\zeta^2}, & i=k\neq j \\ \frac{4w_i}{\zeta^2}, & i\neq j=k \\ 0, & \text{otherwise} \end{cases}.$$

The inverse of Hessian is

$$\nabla^{-2}_{u_i,u_j} f_{\text{gpow}} = -\delta(i,j)\frac{u_i}{\nabla_{u_i} f_{\text{gpow}}} - \frac{4\|w\|^2}{k_2\zeta}\frac{\alpha_i}{\nabla_{u_i} f_{\text{gpow}}}\frac{\alpha_j}{\nabla_{u_j} f_{\text{gpow}}},$$

$$\nabla^{-2}_{u_i,w_j} f_{\text{gpow}} = \frac{2}{k_2}\frac{\alpha_i}{\nabla_{u_i} f_{\text{gpow}}}w_j,$$

$$\nabla^{-2}_{w_i,w_j} f_{\text{gpow}} = \delta(i,j)\frac{\zeta}{2} + \frac{(\zeta^{-1}\varphi + 4k_1)}{k_2}w_iw_j,$$

where

$$k_1 = \left\langle \frac{\alpha}{\nabla_u f}, \frac{\alpha}{u}\right\rangle, \quad k_2 = -\left(1+\frac{\|w\|^2}{\varphi}\right) - \frac{4k_1\|w\|^2}{\zeta}.$$

A.2 Power Mean Cone

Note that $\varphi = \prod_{i\in[\![d]\!]} u_i^{\alpha_i}$, $\zeta = \varphi - w$ and $\tau \in \mathbb{R}^d$ with $\tau_i = \frac{\alpha_i}{u_i\zeta}, \forall i \in [\![d]\!]$. The third-order derivative is

$$
\begin{aligned}
\nabla^3_{u_i,u_j,u_k} f_{\text{powm}} &= -w\varphi\tau_i\tau_j\tau_k(2w+\zeta) + \begin{cases} -\frac{w\varphi\tau_i\tau_k}{u_i} - \frac{2\varphi\tau_i}{u_i^2} - \frac{2}{u_i^3}, & i=j=k \\ -\frac{w\varphi\tau_i\tau_k}{u_i}, & i=j\neq k \\ 0, & \text{otherwise} \end{cases} \\
&\quad + \begin{cases} -\frac{2w\varphi\tau_i\tau_j}{u_i}, & i=j=k \\ -\frac{w\varphi\tau_i\tau_j}{u_i}, & i=k\neq j \\ -\frac{w\varphi\tau_i\tau_j}{u_j}, & i\neq j=k \\ 0, & \text{otherwise} \end{cases}, \\
\nabla^3_{u_i,u_j,w} f_{\text{powm}} &= \varphi\tau_i\tau_j\left(\frac{2\varphi}{\zeta} - 1\right) + \delta(i,j)\frac{\tau_i\varphi}{u_i\zeta}, \\
\nabla^3_{u_i,w,w} f_{\text{powm}} &= -\frac{2\varphi\tau_i}{\zeta^2}, \\
\nabla^3_{w,w,w} f_{\text{powm}} &= \frac{2}{\zeta^3},
\end{aligned}
$$

The inverse of Hessian is

$$
\begin{aligned}
\nabla^{-2}_{u_i,u_j} f_{\text{powm}} &= \delta(i,j)\frac{u_i^2}{s_{0,i}} + \frac{\varphi}{\zeta s_2}\frac{\alpha_i u_i}{s_{0,i}}\frac{\alpha_j u_j}{s_{0,j}}, \\
\nabla^{-2}_{u_i,w} f_{\text{powm}} &= \frac{\varphi}{s_2}\frac{\alpha_i u_i}{s_{0,i}}, \\
\nabla^{-2}_{w,w} f_{\text{powm}} &= \zeta^2 + \frac{s_1}{s_2}\varphi^2,
\end{aligned}
$$

where

$$
\begin{aligned}
s_{0,i} &= 1 + \alpha_i\varphi\zeta^{-1}, \forall i \in [\![d]\!], \\
s_1 &= \sum_{i\in[\![d]\!]} \frac{\alpha_i^2}{s_{0,i}}, \\
s_2 &= 1 - \varphi\zeta^{-1}s_1.
\end{aligned}
$$

A.3 Relative Entropy Cone

For a relative entropy cone, we define $\alpha_i = \frac{1}{\zeta+2w_i}, \forall i \in [\![d]\!]$, $\beta_i = \log\left(\frac{w_i}{v_i}\right), \forall i \in [\![d]\!]$, and $\gamma_i = \sum_{j\in[\![d]\!]\setminus\{i\}} w_j\beta_j, \forall i \in [\![d]\!]$. Then:

$$
\begin{aligned}
\nabla^3_{u,u,u} f &= -\frac{2}{\zeta^3} \\
\nabla^3_{u,u,v_i} f &= -\frac{2\sigma_i}{\zeta^2} \\
\nabla^3_{u,u,w_i} f &= -\frac{2\tau_i}{\zeta^2} \\
\nabla^3_{u,v_i,v_j} f &= -\frac{2\sigma_i\sigma_j}{\zeta} - \delta(i,j)\frac{\sigma_i}{v_i\zeta}, \\
\nabla^3_{u,v_i,w_j} f &= -\frac{2\sigma_i\tau_j}{\zeta} + \delta(i,j)\frac{1}{v_i\zeta^2}, \\
\nabla^3_{u,w_i,w_j} f &= -\frac{2\tau_i\tau_j}{\zeta} - \delta(i,j)\frac{1}{w_i\zeta^2}, \\
\nabla^3_{v_i,v_j,v_k} f &= -2\sigma_i\sigma_j\sigma_k + \begin{cases} -\frac{3\sigma_i^2}{v_i} - 2\frac{\sigma_i}{v_i^2} - \frac{2}{v_i^3} & i=j=k, \\ -\frac{\sigma_i\sigma_k}{v_1} & i=j\neq k, \\ 0 & \text{otherwise}, \end{cases} \\
\nabla^3_{v_i,v_j,w_k} f &= -2\sigma_i\sigma_j\tau_k + \begin{cases} \frac{1}{v_i^2\zeta} - \frac{\tau_i\sigma_i}{v_i} + \frac{2\sigma_i}{v_i\zeta} & i=j=k, \\ -\frac{\tau_i\sigma_i}{v_i} & i=j\neq k, \\ \frac{\sigma_i}{v_j\zeta} & i\neq j=k, \\ 0 & \text{otherwise}, \end{cases} \\
\nabla^3_{v_i,w_j,w_k} f &= -2\sigma_i\tau_j\tau_k + \begin{cases} 2\frac{\tau_k}{v_i\zeta} - \frac{1}{v_i\zeta^2} & i=j=k, \\ \frac{\tau_k}{v_i\zeta} & i=j\neq k, \\ -\frac{\sigma_i}{w_j\zeta} & i\neq j=k, \\ 0 & \text{otherwise}, \end{cases} \\
\nabla^3_{w_i,w_j,w_k} f &= -2\tau_i\tau_j\tau_k + \begin{cases} -\frac{3\tau_k}{w_i\zeta} - \frac{1}{w_i^2\zeta} - \frac{2}{w_i^3} & i=j=k, \\ -\frac{\tau_k}{w_i\zeta} & i=j\neq k, \\ 0 & \text{otherwise}. \end{cases}
\end{aligned}
$$

The inverse Hessian is

$$
\begin{aligned}
\nabla_{u,u}^{-2} f &= \zeta^2 - \sum_{i \in [\![d]\!]} \alpha_i w_i \left(w_i \beta_i - \zeta + \tau_i \zeta \left(\zeta + \zeta \beta_i + w_i \beta_i \right) \right), \\
\nabla_{u,v_i}^{-2} f &= -v_i w_i \alpha_i \left(u - \gamma_i - 2 w_i \beta_i \right), \\
\nabla_{u}^{-2} w_i f &= -w_i^2 \alpha_i \left(\beta_i \zeta + \gamma_i \right), \\
\nabla_{v_i,v_j}^{-2} f &= \delta(i,j) v_i^2 \alpha_i \left(\zeta + w_i \right), \\
\nabla_{v_i,w_j}^{-2} f &= \delta(i,j) v_i w_i^2 \alpha_i, \\
\nabla_{w_i,w_j}^{-2} f &= \delta(i,j) w_i^2 \alpha_i \left(\zeta + w_i \right).
\end{aligned}
$$

Acknowledgements Yuwen Chen is financially supported by the Clarendon Scholarship from University of Oxford. We also would like to thank the anonymous reviewers for their insightful comments and suggestions, which have greatly improved the quality of this paper.

References

1. Andersen, M.S., Dahl, J., Vandenberghe, L.: Implementation of nonsymmetric interior-point methods for linear optimization over sparse matrix cones. Math. Progr. Comput. **2**(3), 167–201 (2010). https://doi.org/10.1007/s12532-010-0016-2
2. Boyd, S., Vandenberghe, L.: Convex Optimization. Cambridge University Press, Cambridge (2004)
3. Boyd, S., El Ghaoui, L., Feron, E., Balakrishnan, V.: Linear Matrix Inequalities in System and Control Theory. Society for Industrial and Applied Mathematics, Philadelphia (1994). https://doi.org/10.1137/1.9781611970777
4. Chandrasekaran, V., Shah, P.: Relative entropy optimization and its applications. Math. Program. **161**(1), 1–32 (2017). https://doi.org/10.1007/s10107-016-0998-2
5. Chares, P.R.: Cones and interior-point algorithms for structured convex optimization involving powers and exponentials. PhD thesis, Université catholique de Louvain (2009)
6. Coey, C., Kapelevich, L., Vielma, J.P.: Solving natural conic formulations with Hypatia.jl. INFORMS J. Comput. **34**(5), 2686–2699 (2022). https://doi.org/10.1287/ijoc.2022.1202
7. Coey, C., Kapelevich, L., Vielma, J.P.: Performance enhancements for a generic conic interior point algorithm. Math. Program. Comput. **15**(1), 53–101 (2023). https://doi.org/10.1007/s12532-022-00226-0
8. Cornuéjols, G., Peña, J., Tütüncü, R.: Optimization Methods in Finance, 2nd edn. Cambridge University Press, Cambridge (2018). https://doi.org/10.1017/9781107297340
9. Cuturi, M., Doucet, A.: Fast computation of Wasserstein barycenters. In: Xing, E.P., Jebara, T. (eds.) Proceedings of the 31st International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 32, pp. 685–693. PMLR, Bejing, China (2014)
10. Dahl, J., Andersen, E.D.: A primal-dual interior-point algorithm for nonsymmetric exponential-cone optimization. Math. Program. (2021). https://doi.org/10.1007/s10107-021-01631-4

11. Davis, T.A.: Algorithm 849: a concise sparse cholesky factorization package. ACM Trans. Math. Softw. **31**(4), 587–591 (2005). https://doi.org/10.1145/1114268.1114277
12. Domahidi, A., Chu, E., Boyd, S.: ECOS: an SOCP solver for embedded systems. In: 2013 European Control Conference (ECC), pp. 3071–3076. IEEE, Zurich (2013). https://doi.org/10.23919/ECC.2013.6669541
13. Garstka, M., Cannon, M., Goulart, P.: COSMO: a conic operator splitting method for convex conic problems. J. Optim. Theory Appl. **190**(3), 779–810 (2021)
14. Goulart, P.J., Chen, Y.: Clarabel: an interior-point solver for conic programs with quadratic objectives. arXiv. arXiv:2405.12762 (2024)
15. Kapelevich, L., Andersen, E.D., Vielma, J.P.: Computing conjugate barrier information for nonsymmetric cones. J. Optim. Theory Appl. (2022). https://doi.org/10.1007/s10957-022-02076-1
16. Karimi, M., Tunçel, L.: Domain-driven solver (DDS) version 2.1: A MATLAB-based software package for convex optimization problems in domain-driven form. Math. Programm. Comput. (2023). https://doi.org/10.1007/s12532-023-00248-2
17. Luo, Z.-Q., Yu, W.: An introduction to convex optimization for communications and signal processing. IEEE J. Sel. Areas Commun. **24**(8), 1426–1438 (2006). https://doi.org/10.1109/JSAC.2006.879347
18. Mehrotra, S.: On the implementation of a primal-dual interior point method. SIAM J. Optim. **2**(4), 575–601 (1992). https://doi.org/10.1137/0802028
19. MOSEK APS.: MOSEK Modeling Cookbook. (2023). https://docs.mosek.com/MOSEKModelingCookbook-a4paper.pdf
20. Murray, R., Chandrasekaran, V., Wierman, A.: Signomial and polynomial optimization via relative entropy and partial dualization. Math. Program. Comput. **13**(2), 257–295 (2021). https://doi.org/10.1007/s12532-020-00193-4
21. Nesterov, Y.E., Todd, M.J.: Self-scaled barriers and interior-point methods for convex programming. Math. Oper. Res. **22**(1), 1–42 (1997)
22. Nesterov, Y.E., Todd, M.J.: Primal-dual interior-point methods for self-scaled cones. SIAM J. Optim. **8**(2), 324–364 (1998). https://doi.org/10.1137/S1052623495290209
23. Nesterov, Y.: Towards non-symmetric conic optimization. Opt. Methods Softw. **27**(4–5), 893–917 (2012). https://doi.org/10.1080/10556788.2011.567270
24. Nocedal, J., Wright, S.: Numerical Optimization. Springer Series in Operations Research and Financial Engineering. Springer, New York, NY (2006)
25. O'Donoghue, B., Chu, E., Parikh, N., Boyd, S.: Conic optimization via operator splitting and homogeneous self-dual embedding. J. Optim. Theory Appl. **169**(3), 1042–1068 (2016). https://doi.org/10.1007/s10957-016-0892-3
26. Papp, D., Yildiz, S.: Sum-of-squares optimization without semidefinite programming. SIAM J. Optim. **29**(1), 822–851 (2019). https://doi.org/10.1137/17M1160124
27. Roy, S., Xiao, L.: On self-concordant barriers for generalized power cones. Opt. Lett. **16**(2), 681–694 (2022). https://doi.org/10.1007/s11590-021-01748-7
28. Serrano, S.A.: Algorithms for unsymmetric cone optimization and an implementation for problems with the exponential cone. PhD thesis, Stanford University (2015)
29. Skajaa, A., Ye, Y.: A homogeneous interior-point algorithm for nonsymmetric convex conic optimization. Math. Program. **150**(2), 391–422 (2015). https://doi.org/10.1007/s10107-014-0773-1
30. Terlaky, T., Vial, J.P.: Computing maximum likelihood estimators of convex density functions. SIAM J. Sci. Comput. **19**(2), 675–694 (1998). https://doi.org/10.1137/S1064827595286578
31. Tunçel, L.: Generalization of primal-dual interior-point methods to convex optimization problems in conic form. Found. Comput. Math. **1**(3), 229–254 (2001). https://doi.org/10.1007/s002080010009
32. Vandenberghe, L.: The CVXOPT Linear and Quadratic Cone Program Solvers. (2010)
33. Vanderbei, R.J.: Symmetric quasidefinite matrices. SIAM J. Optim. **5**(1), 100–113 (1995). https://doi.org/10.1137/0805005
34. Ye, Y., Todd, M.J., Mizuno, S.: An O($\sqrt{nL}$)-iteration homogeneous and self-dual linear programming algorithm. Math. Oper. Res. **19**(1), 53–67 (1994). https://doi.org/10.1287/moor.19.1.53