Interior-Point Methods for Constrained Quadratic Programming: A Theoretical and Computational Study

Marco Tamberi - Vincenzo Rocchi

November 6, 2025

Contents

1 Introduction

This report aims to analyze the use of interior-point methods (IPMs) for solving constrained convex quadratic programs (QPs) with polyhedral feasible regions. The structure is simple: first, we provide a clear theoretical motivation for why applying IPMs to this specific class of problems is the most suitable choice. Then, we move to a practical perspective, evaluating their performance on representative instances and comparing the results with those obtained using a general-purpose QP solver.

Constrained QPs arise in numerous application areas. In finance, portfolio models balance expected return and risk under budget and regulatory constraints. In agriculture and operations research, planning models allocate limited resources to maximize yields or profits. In machine learning, support vector machines (SVMs) are trained by solving a convex QP with linear constraints, where the separating hyperplane corresponds to the solution of such a program.

In our specific setting, **Project 34**, the feasible region imposes *simplex* constraints blockwise. This situation is common whenever decision variables encode probabilities, convex weights, or categorical assignments. The structure naturally suggests the use of IPMs, and in particular motivates a closer look at primal–dual feasible-start variants.

1.1 An Overview of the Function

We formalize the optimization problem at the core of the project. Consider the quadratic program **(P)** defined over a set K of disjoint simplices:

$$\min \left\{ x^\top Q x + q^\top x \;:\; \sum_{i \in I^k} x_i = 1 \;\; \forall\, k \in K, \;\; x \geq 0 \right\},$$

where $x \in \mathbb{R}^n$, the index sets I^k form a partition of $\{1, \ldots, n\}$, and $Q \in \mathbb{R}^{n \times n}$ is symmetric positive semidefinite.

[Formally, the partition conditions are $\bigcup_{k \in K} I^k = \{1, \ldots, n\}$ and $I^h \cap I^k = \varnothing$ for all h, k. [write this if to include as i.e., after partition of [1,...,n]]

Notation. We adopt the standard convention for the quadratic objective

$$\tfrac{1}{2} x^\top Q x + q^\top x,$$

where the factor $\frac{1}{2}$ simplifies derivatives: for symmetric Q, $\nabla(\frac{1}{2} x^\top Q x) = Qx$ and $\nabla^2(\frac{1}{2} x^\top Q x) = Q$. Whenever the shorthand "qx" is used, it denotes the inner product $q^\top x$.

[At first glance, the model may seem intimidating, but we can unpack it step by step. The first observation is that it is a nonhomogeneous quadratic function. The term $x^\top Q x$ arises by taking a square matrix Q and multiplying

the variable vector x on the left and on the right:

$$f(x) \;=\; \tfrac{1}{2}\, x^\top Q x \;=\; \tfrac{1}{2}\left[\underbrace{\sum_{i=1}^{n} Q_{ii}\, x_i^2}_{\text{elements on the diagonal}} \;+\; \underbrace{\sum_{i=1}^{n} \sum_{\substack{j=1 \\ j\neq i}}^{n} Q_{ij}\, x_i x_j}_{\text{elements outside the diagonal}} \right], \qquad Q \in \mathbb{R}^{n\times n}.$$

]

The function also includes the linear term $q^\top x$, with $q \in \mathbb{R}^n$. Since Q is positive semidefinite, the quadratic form $\frac{1}{2}x^\top Qx$ is convex, which allows us to state that (P) is a convex quadratic program.

[Formally, if $Q \succeq 0$ (Q is positive semidefinite) then all eigenvalues $\lambda \geq 0$. Consequently, the quadratic form $d^\top Qd$ is nonnegative for every $d \in \mathbb{R}^n$, i.e., $d^\top Qd \geq 0$ for all $d \in \mathbb{R}^n$.]

The curvature of the quadratic form depends on the eigenvalues of Q: larger eigenvalues correspond to steeper curvature in the associated eigen-directions. If $Q \succ 0$, the function is strictly convex and has a unique minimum (in the homogeneous case) at $x = 0$. If $Q \succeq 0$ but singular, the function may be flat along some directions, but it remains convex.

[Geometrically, the quadratic "opens upward" in every eigen-direction, and is completely flat along eigenvectors corresponding to zero eigenvalues. For a homogeneous quadratic with $Q \succeq 0$, the minimum value is 0 (but the function is unbounded above).]

Including the linear term $q^\top x$ modifies the previous picture. If $Q \succ 0$, the problem remains strictly convex and admits a unique minimum. If $Q \succeq 0$, a finite minimum exists if and only if $q \in \mathrm{Im}(Q)$, equivalently $q \perp \ker(Q)$. In this case, the quadratic term alone attains its minimum value 0 on $\ker(Q)$, but the addition of the linear term may destroy boundedness unless q is flat along $\ker(Q)$.

[Along eigenvectors corresponding to zero eigenvalues, the quadratic term is flat. In such directions, the objective reduces to a purely linear function: if the linear component is nonzero, the problem is unbounded below; if it vanishes, a minimizer may still exist.]

The domain is defined by linear equalities and nonnegativity constraints. The indices are divided into disjoint subsets I^k, each representing a block of variables subject to

$$\sum_{i\in I^k} x_i = 1, \qquad x_i \geq 0.$$

In other words, each block of variables lies on a simplex.

[Equivalently, each block must satisfy two conditions: all variables are non-negative and their sum is exactly 1.]

Geometrically, each block I^k defines a standard simplex. Since the blocks

are disjoint, the overall domain is the Cartesian product of these simplices:

$$\mathcal{X} \;=\; \prod_{k\in K} \Delta^{|I^k|-1},$$

where $\Delta^{|I^k|-1}$ denotes the standard simplex of dimension $|I^k|-1$.

[Intuitively, this means that each group of variables cannot vary freely in $\mathbb{R}^n$, but is confined within the simplex polytope.]

The feasible region is thus a closed, bounded (and therefore compact) convex polytope. Interior points correspond to strictly positive vectors ($x_i > 0$), while vertices represent "pure" choices in which, for each block, a single coordinate equals 1 and the others are 0.

2 A Brief Preamble to Interior-Point Methods

Before introducing interior point methods, we need to provide some theoretical background to better understand the context and how these methods work. A standard convex quadratic program can be written as

$$\min\left\{\, \tfrac{1}{2}x^\top Qx + q^\top x \;:\; Ax \le b \,\right\},$$

with $Q \succeq 0$. This is referred to as the *primal* problem. In many cases, analysis is simplified by considering the associated *dual* problem.

[

Consistency with our model. In our specific instance (product of simplices) the constraints take the form $Ex = \mathbf{1}$ and $x \ge 0$; we will switch to this equalities+nonnegativity form when writing the KKT system for our algorithm.]

2.1 Dual Methods

The starting point is the Lagrangian function:

$$L(x,\lambda) \;=\; \tfrac{1}{2}x^\top Qx + q^\top x \;+\; \lambda^\top(Ax-b), \qquad \lambda \ge 0,$$

where the multipliers $\lambda \in \mathbb{R}^m$ penalize constraint violations. The associated dual function is

$$\psi(\lambda) \;=\; \inf_{x\in\mathbb{R}^n} L(x,\lambda),$$

Which provides a lower bound to the optimal value $\nu(P)$ of the primal problem. The dual problem is therefore

$$(D) \qquad \max_{\lambda\ge 0} \; \psi(\lambda).$$

[Equivalently, one can write the Lagrangian as $\frac{1}{2}x^\top Qx + q^\top x + \lambda^\top(b-Ax)$, with the opposite sign convention for λ; both forms are standard and lead to the same dual.]

The dual perspective is powerful: under convexity assumptions and a constraint qualification such as Slater's condition (existence of a strictly feasible point), *strong duality* holds:

$$\nu(D) \;=\; \nu(P),$$

So the optimal value of the dual coincides with that of the primal. This enables constrained problems to be solved indirectly through their dual, which may in some cases be simpler or better structured.

However, the dual approach also presents some limitations. The dual function $\psi(\lambda)$ is concave but may be nonsmooth (not even C^1 if the primal objective is not strictly convex), and the primal vector $x(\lambda)$ is typically infeasible until the end of the iterations. In other words, dual methods provide valid lower bounds early, but recovering a feasible primal solution is less immediate.

[

Remark (specialization to our constraints). For later use, when $Ex = \mathbf{1}$ and $x \geq 0$, a convenient Lagrangian is

$$L(x, y, z) \;=\; \tfrac{1}{2} x^\top Q x + q^\top x \;+\; y^\top (Ex - \mathbf{1}) \;-\; z^\top x, \qquad z \geq 0,$$

with y free (equalities) and z the multipliers for nonnegativity.]

Algorithmic Desiderata. These limitations incentive methods that:

1. yields an unconstrained dual optimization problem,
2. operate with a smooth (preferably C^2) dual function ψ,
3. maintain (strict) primal feasibility throughout the iterations.

2.2 Barrier Methods

To address the limitations of dual methods, we now look at *barrier methods*. These replace explicit inequality constraints with penalization terms that diverge as iterates approach the boundary of the feasible region. A natural choice is the *logarithmic barrier*:

$$f_\gamma(x) \;=\; f(x) \;-\; \gamma \sum_{i=1}^{m} \ln\big(b_i - A_i x\big), \qquad \gamma > 0,$$

where γ controls the strength of the penalization.

This transformation ensures that $f_\gamma(x)$ is infinite for infeasible points ($x \notin X$) or boundary points ($x \in \partial X$), so all iterates remain strictly feasible. Moreover, on the interior of X, f_γ is C^2 and, if f is convex, then f_γ is strictly convex, enabling an effective use of Newton's method.

A key concept is the *central path*: as the barrier parameter γ varies, we obtain a family of problems (P_γ) with solutions x_γ that converge to the optimum $x^\star$ of the original problem. In particular,

$$\gamma \to \infty \;\Rightarrow\; x_\gamma \to \text{analytic center of } X, \qquad \gamma \to 0 \;\Rightarrow\; x_\gamma \to x^\star.$$

This continuous trajectory, the central path, serves as a stable guide for the iterates.

From a computational perspective, barrier methods have notable advantages. Newton's method applied to f_γ converges very rapidly when initialized near the central path, often achieving superlinear or even quadratic convergence.

Iteration complexity. Classical analyses give an iteration bound of

$$O\big(m \log(1/\varepsilon)\big),$$

where m is the number of inequality constraints. Using the self-concordant barrier framework, this improves to

$$O\big(\sqrt{\nu}\, \log(1/\varepsilon)\big),$$

where ν is the barrier parameter (equal to the number of barrier terms in the logarithmic case). The dominant cost per iteration lies in solving the Newton linear systems.

[In the generic form $Ax \le b$, we have $\nu = m$. In our formulation $Ex = \mathbf{1}$, $x \ge 0$, the barrier applies only to the n nonnegativity constraints, so $\nu = n$. For dense problems, solving the Newton systems typically costs at least $O(n^3)$ per iteration.]

[

Remark (link to our constraints). For $Ex = \mathbf{1}$ and $x > 0$, the barrier term specializes to $-\gamma \sum_{i=1}^{n} \ln x_i$, while the equalities are handled by Lagrange multipliers. In the primal–dual setting introduced next, the target complementarity parameter μ coincides with γ along the central path.]

3 Interior-Point Methods

As we have seen, barrier methods transform a constrained problem into a smooth unconstrained one on the interior of the feasible set, ensuring strict feasibility and strenghtening the use of Newton's method. This makes them among the most effective and widely used approaches for solving Linear Programming (LP), Quadratic Programming (QP), and Semidefinite Programming (SDP) problems. Their most powerful incarnation is in *primal–dual* formulations, where the primal and dual problems are solved simultaneously, keeping both nearly consistent throughout the iterations.

3.1 KKT Conditions and Relaxation

Consider the constrained quadratic program

$$(P) \qquad \min \ \tfrac{1}{2} x^\top Q x + q^\top x \quad \text{s.t.} \quad Ax \le b,$$

which can be reformulated by introducing slack variables $s \ge 0$ [(a standard trick to turn inequalities into equalities)] as

$$Ax + s = b, \qquad Qx + A^\top \lambda + q = 0, \qquad \lambda \ge 0,\ s \ge 0, \qquad \lambda_i s_i = 0.$$

These are precisely the Karush–Kuhn–Tucker (KKT) conditions:

1. primal feasibility (KKT-F): $Ax + s = b$, $s \ge 0$;
2. dual feasibility (KKT-G): $Qx + A^\top \lambda + q = 0$, $\lambda \ge 0$;
3. complementarity (KKT-CS): $\lambda_i s_i = 0$ for $i = 1, \dots, m$.

The primal–dual interior-point method replaces the complementarity condition $\lambda_i s_i = 0$ with a relaxed version:

$$\lambda_i s_i \ = \ \mu, \qquad i = 1, \dots, m,$$

with $\mu > 0$ gradually decreased to zero. Rather than enforcing complementarity exactly [(which is numerically challenging)], this approach follows a *central path* that connects the solutions of the perturbed problems (P_μ) to the optimal solution of the original problem. For the logarithmic barrier [along the central path] with weight γ, the relation is $\mu = \gamma$.

3.2 Newton Step and Computational Aspects

At the heart of primal–dual interior-point methods is the Newton step applied to the relaxed KKT system. By linearizing the nonlinear complementarity conditions, we obtain a linear system whose solution gives the search direction for the next iterate. The main computational effort in each iteration is solving this system, which usually requires heavy matrix factorizations.

[Explicitly, introducing $\Lambda = \operatorname{diag}(\lambda)$ and $S = \operatorname{diag}(s)$, the complementarity condition is written as $\Lambda S u = \mu u$. Applying Newton's method gives the block linear system

$$\begin{bmatrix} Q & A^\top & 0 \\ A & 0 & I \\ 0 & S & \Lambda \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta \lambda \\ \Delta s \end{bmatrix} = - \begin{bmatrix} r_D \\ r_P \\ r_C \end{bmatrix},$$

with residuals $r_D = Qx + A^\top \lambda + q$, $r_P = Ax + s - b$, and $r_C = \Lambda S u - \mu u$. For dense problems this entails a worst-case cost of $O(n^3)$ per iteration. To improve efficiency and stability, predictor–corrector schemes are often used: an affine-scaling predictor provides a descent direction, then a correction step reintroduces the centering term to follow the central path more closely. This reduces the number of iterations and accelerates convergence in practice.]

Specialization. In our problem, the constraints take the form of equalities and nonnegativity and the general interior-point framework adapts directly to this structure.

[

Specialization to our constraints ($Ex = \mathbf{1},\ x \geq 0$). For Project 34 the constraints are equalities plus nonnegativity. Using the Lagrangian

$$L(x, y, z) = \tfrac{1}{2} x^\top Q x + q^\top x + y^\top (Ex - \mathbf{1}) - z^\top x, \qquad z \geq 0,$$

the KKT conditions become

$$\begin{aligned} &\text{primal feasibility:} && Ex = \mathbf{1}, \quad x \geq 0, \\ &\text{dual feasibility:} && Qx + q + E^\top y - z = 0, \quad z \geq 0, \\ &\text{complementarity:} && x_i z_i = 0, \quad i = 1, \ldots, n. \end{aligned}$$

Here z plays the role of the dual multipliers for $x \geq 0$ (analogous to λ for inequalities in the generic case). In a primal–dual interior-point method, complementarity is relaxed to $x_i z_i = \mu$ with $\mu > 0$ decreased to zero (on the central path for a log barrier of weight γ, one has $\mu = \gamma$). Let $X = \operatorname{diag}(x)$ and $Z = \operatorname{diag}(z)$, and define residuals

$$r_D = Qx + q + E^\top y - z, \qquad r_P = Ex - \mathbf{1}, \qquad r_C = XZ\mathbf{1} - \mu\mathbf{1}.$$

Linearizing the KKT system yields the Newton equations

$$\begin{bmatrix} Q & E^\top & -I \\ E & 0 & 0 \\ Z & 0 & X \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \\ \Delta z \end{bmatrix} = - \begin{bmatrix} r_D \\ r_P \\ r_C \end{bmatrix}.$$

]

Elimination and reduced system. The Newton equations can be simplified through block elimination, yielding a Schur-complement system in $(\Delta x, \Delta y)$ which make it more efficient to solve in practice.

[For a **feasible-start** primal–dual interior-point method we assume that the iterates always satisfy primal and dual feasibility, $Ex = \mathbf{1}$ and $Qx + E^\top y - z + q = 0$, so that only the complementarity condition is relaxed. From the third block row of the Newton system we obtain

$$X\,\Delta z = -\,r_C - Z\,\Delta x, \qquad r_C = XZ\mathbf{1} - \mu\mathbf{1}.$$

Hence

$$\Delta z = -\,X^{-1} r_C - X^{-1} Z\,\Delta x.$$

Substituting into the first block row $Q\Delta x + E^\top \Delta y - \Delta z = -\,r_D$ gives the reduced (Schur–complement) system in $(\Delta x, \Delta y)$:

$$\boxed{(Q + X^{-1}Z)\,\Delta x + E^\top \Delta y = -\,r_D - X^{-1} r_C}, \qquad E\,\Delta x = -\,r_P.$$

Because a feasible-start method keeps both primal and dual constraints exactly satisfied at each iteration, the residuals r_P and r_D vanish identically, and only the complementarity residual r_C remains to be corrected. Therefore, the linearized system simplifies further to

$$\boxed{(Q + X^{-1}Z)\,\Delta x + E^\top \Delta y = -\,X^{-1} r_C, \qquad E\,\Delta x = \mathbf{0}.}$$

This form corrects only complementarity errors while maintaining primal and dual feasibility at every iteration.]

[

Step sizes and update. Choose primal and dual step lengths by the fraction-to-the-boundary rule:

$$\alpha_{\text{pri}} = \min\Big(1,\ 0.99\cdot \min_{i:\,\Delta x_i<0} -\tfrac{x_i}{\Delta x_i}\Big), \qquad \alpha_{\text{dual}} = \min\Big(1,\ 0.99\cdot \min_{i:\,\Delta z_i<0} -\tfrac{z_i}{\Delta z_i}\Big),$$

and set $\alpha = \min(\alpha_{\text{pri}}, \alpha_{\text{dual}})$ to ensure $x + \alpha\Delta x > 0$ and $z + \alpha\Delta z > 0$.

Variable updates.

$$x \leftarrow x + \alpha\Delta x, \qquad y \leftarrow y + \alpha\Delta y, \qquad z \leftarrow z + \alpha\Delta z.$$

Complementarity update. After each iteration, update the complementarity measure as

$$\mu = \frac{x^\top z}{n}.$$

On the central path, each component satisfies $x_i z_i = \mu$, so the average relation holds exactly. Because our feasible-start method maintains primal and dual feasibility, μ reflects only the remaining *duality gap* between primal and dual objectives. Since the iterates are not exactly on the central path, the next target value is set as

$$\mu_{\text{target}} = \sigma\,\frac{x^\top z}{n}, \qquad \sigma \in (0, 1).$$

The parameter σ controls the centering: values close to 1 give safer but slower progress, while smaller values produce more aggressive steps that close the duality gap faster. In our implementation we would typically take $\sigma \in [0.1, 0.5]$.
]

Strengths and Limitations Primal–dual interior-point methods have several fundamental strengths:

- they maintain primal and dual feasibility (or at least remain very close to it) at each iteration;

- they provide accurate control of the duality gap through the relaxed complementarity condition; [$\lambda_i s_i = \mu$ generic inequality form; in our model with $x \geq 0$, the relaxed complementarity condition reads $x_i z_i = \mu$.] ;
- they achieve polynomial-time convergence, with iteration complexity bounded by $O\big(\sqrt{\nu}\,\log(1/\varepsilon)\big)$;
- they are broadly applicable across LP, QP, SOCP, and SDP problems.

The main drawback is the per-iteration cost of solving large linear systems. In practice, however, this is compensated by the excellent numerical stability of the method and by the relatively small number of iterations usually required.

3.3 Pseudocodes - *NEW PARAGRAPH*

The previous subsections derived the feasible-start primal–dual interior-point framework for our quadratic program, highlighting how primal and dual feasibility are preserved while complementarity is progressively reduced. Having established the theoretical formulation and the simplified Newton system, we can now translate these steps into an explicit computational procedure.

From theory to implementation. Among the possible algorithmic variants, the two main choices are the *feasible-start* and the *infeasible-start* methods. The latter allows initialization from arbitrary (possibly infeasible) points, but in our case the feasible region is a compact polyhedron, a Cartesian product of simplices, so a strictly feasible starting point can be obtained without special preprocessing; for instance by assigning uniform positive values within each simplex block. Hence, adopting the feasible-start variant provides conceptual clarity and numerical stability.

[In practice, small numerical errors may cause slight deviations from exact feasibility; these are corrected implicitly through each Newton step and monitored via the residuals r_P and r_D.]

The algorithms presented below implement exactly this formulation. The first uses a fixed centering parameter σ, while the second incorporates Mehrotra's predictor–corrector strategy, which adaptively adjusts σ at each iteration to accelerate convergence.

[Throughout the algorithms, $X = \mathrm{diag}(x)$ and $Z = \mathrm{diag}(z)$ denote diagonal matrices of primal and dual variables, respectively.]

[In the exact feasible-start formulation, the primal and dual constraints remain satisfied at all iterations, so $r_P = r_D = 0$ analytically. In implementation, however, numerical errors may cause small residuals, and it is standard to monitor $\|r_P\|_\infty$ and $\|r_D\|_\infty$ alongside μ in the stopping criterion.]

Starting from the standard algorithm:

Algorithm 1: Feasible-Start Primal–Dual Interior-Point Method (Fixed Centering Parameter)

Input: $Q \succeq 0$, q, block-equality matrix E, tolerances ϵ_{feas}, ϵ_{comp}, centering parameter $\sigma \in (0, 0.5)$, maximum iterations *max_iter*

Output: Approximate optimizer (x, y, z) satisfying $Ex = \mathbf{1}$, $x > 0$, $z > 0$, $Qx + q + E^\top y - z \approx 0$

Step 1: Initialization.;

For each block I^k: set $x_i \leftarrow 1/|I^k|$ for $i \in I^k$; `// Ex = 1, x > 0`

$w \leftarrow Qx + q$; choose $\delta > 0$ and set $y_k \leftarrow -\min_{i \in I^k} w_i + \delta$ for all k;;

$z \leftarrow w + E^\top y$; `// z > 0 by construction`

$\mu \leftarrow (x^\top z)/n$.;

Step 2: Iterative loop.;

for $k = 1$ **to** *max_iter* **do**

  Step 2.1: Residual computation.;

  $r_D \leftarrow Qx + q + E^\top y - z$; $r_P \leftarrow Ex - \mathbf{1}$; $r_C \leftarrow XZ\mathbf{1} - \mu\mathbf{1}$.;

  Step 2.2: System assembly.;

  $X \leftarrow \text{diag}(x)$, $Z \leftarrow \text{diag}(z)$, $H \leftarrow Q + X^{-1}Z + \delta I$;

  `// regularized`

  $\tilde{\mu} \leftarrow \sigma\,\mu$; $\text{rhs} \leftarrow \tilde{\mu}\,X^{-1}\mathbf{1} - Z\mathbf{1}$.;

  $S \leftarrow EH^{-1}E^\top$; $b \leftarrow EH^{-1}\text{rhs}$.;

  Step 2.3: Direction computation.;

  Solve $S\,\Delta y = b$; $\Delta x \leftarrow H^{-1}(\text{rhs} - E^\top \Delta y)$; $\Delta z \leftarrow Q\,\Delta x + E^\top \Delta y$.;

  Step 2.4: Step-size selection.;

  $\alpha_{\text{pri}} \leftarrow \min\Big(1,\, 0.99 \cdot \min_{i:\,\Delta x_i < 0} -x_i/\Delta x_i\Big)$;;

  $\alpha_{\text{dual}} \leftarrow \min\Big(1,\, 0.99 \cdot \min_{i:\,\Delta z_i < 0} -z_i/\Delta z_i\Big)$;;

  $\alpha \leftarrow \min(\alpha_{\text{pri}}, \alpha_{\text{dual}})$.;

  Step 2.5: Update.;

  $x \leftarrow x + \alpha\,\Delta x$; $y \leftarrow y + \alpha\,\Delta y$; $z \leftarrow z + \alpha\,\Delta z$; $\mu \leftarrow (x^\top z)/n$.;

  Step 2.6: Stopping criterion.;

  if $\|r_P\|_\infty \le \epsilon_{\text{feas}}$ ***and*** $\|r_D\|_\infty \le \epsilon_{\text{feas}}$ ***and*** $\mu \le \epsilon_{\text{comp}}$ **then**

    break

[REMEMBER TO DO! Algorithm 1 doesn't change structurally; you just tighten Step 1 and Step 2.3 to reflect the δ start and the M-based Schur solve.]

Choice of $\delta > 0$ and invertibility of the Schur matrix. A correct initialization and a well-conditioned Newton system require both a strictly interior starting point and a positive definite Schur complement. We first construct a feasible and strictly interior triplet (x^0, y^0, z^0) ensuring $Ex^0 = \mathbf{1}$, $x^0 > 0$, $z^0 > 0$, and $Qx^0 + q + E^\top y^0 - z^0 = 0$. For each block I_k, we set

$$x_i^0 = \frac{1}{|I_k|}, \qquad i \in I_k,$$

and define $w := Qx^0 + q$. Then, for each block k,

$$y_k^0 := -\min_{i \in I_k} w_i + \delta, \qquad \delta > 0,$$

and consequently

$$z_i^0 := w_i + y_k^0, \qquad i \in I_k.$$

By construction $z_i^0 \geq \delta$ for all i, hence $z^0 \geq \delta \mathbf{1} > 0$, and the starting point is strictly interior.

Practical rule for δ. A convenient adaptive choice is

$$\delta = \max\Big\{\varepsilon,\ \tau \cdot \mathrm{range}(w_{I_k})\Big\}, \qquad \varepsilon \in [10^{-10}, 10^{-6}],\ \tau \in [10^{-3}, 10^{-1}],$$

where $\mathrm{range}(w_{I_k}) = \max_{i \in I_k} w_i - \min_{i \in I_k} w_i$. This avoids z_i^0 that are too small (which would cause numerical instability) while maintaining proximity to the initial near-centrality. The initial duality gap satisfies

$$\mu^0 = \frac{(x^0)^\top z^0}{n} \geq \frac{\delta}{n} \sum_i x_i^0 = \frac{\delta}{n} |K|,$$

providing a controlled and finite starting gap.

Invertibility of the Schur matrix via the M formulation. With $X = \mathrm{diag}(x)$ and $Z = \mathrm{diag}(z)$, assuming $x > 0$, $z > 0$, we define

$$M := Z + XQ \qquad (\text{with } Q \succeq 0).$$

Matrix M is symmetric positive definite since, for any $v \neq 0$,

$$v^\top M v = v^\top Z v + v^\top XQv \geq \min_i z_i \, \|v\|^2 > 0.$$

Hence M^{-1} exists and is SPD. The Schur complement associated with the block constraints is

$$S := E\, M^{-1} X\, E^\top.$$

Because $X \succ 0$ and $M^{-1} \succ 0$, also $M^{-1}X \succ 0$. If the rows of E are linearly independent (one per block, summing to 1 on each I_k), then for every nonzero u in the multiplier space,

$$u^\top S u = (E^\top u)^\top (M^{-1}X)(E^\top u) > 0,$$

which implies $S \succ 0$ and therefore invertible. In practice S is small (dimension equal to the number of blocks) and well conditioned as long as z is not too close to zero. When required, a mild regularization can be introduced:

$$M \leftarrow M + \tau I, \qquad \tau \in [10^{-12}, 10^{-8}]\, \|M\|_2,$$

which improves conditioning without changing the theoretical properties.

Operational solution of $S\Delta y = b$ (step 2.3). Forming $S = EM^{-1}XE^\top$ explicitly is discouraged. Instead, its *operator form* is used for products $v \mapsto Sv$:

$$Sv = E\big(M^{-1}(X(E^\top v))\big).$$

This requires: (i) applying $E^\top$; (ii) solving $Mu = X(E^\top v)$; (iii) applying E. Implementation proceeds as follows:

1. **Factorization of M.** Since $M = Z + XQ \succ 0$, a Cholesky or LDL$^\top$ factorization (sparse if Q is sparse) is used and reused for all $M^{-1}(\cdot)$ operations within the iteration.

2. **Solver for S.** If the number of blocks is moderate, $S\Delta y = b$ is solved via explicit Cholesky on assembled S; otherwise, a *preconditioned conjugate-gradient* (CG) method with diagonal or block-Jacobi preconditioner is used, where the product with S follows the operator form above.

3. **Back-substitution.** Once Δy is obtained, compute

   $$\Delta x = M^{-1}\big(-r_c - Xr_d - XE^\top \Delta y\big), \qquad \Delta z = Q\Delta x + E^\top \Delta y + r_d,$$

   for the predictor and corrector variants (with or without Mehrotra correction).

This procedure stabilizes the Newton step, avoids forming a dense Schur matrix, and reduces computational cost when Q is structured.

Alternative H formulation and role of δ. A mathematically equivalent form that emphasizes the explicit regularization parameter is

$$H := Q + X^{-1}Z + \delta I, \qquad (Q \succeq 0,\ \delta > 0),$$

with Schur complement

$$S := EH^{-1}E^\top.$$

For any $v \neq 0$,

$$v^\top Hv = v^\top Qv + v^\top X^{-1}Zv + \delta\|v\|^2 \geq \delta\|v\|^2 > 0,$$

so $H \succ 0$ and, because E has full row rank ($|K|$ independent rows), $S \succ 0$ and invertible. The H formulation thus makes the influence of $\delta > 0$ explicit as a theoretical guarantee of strict positive definiteness and invertibility, while the M formulation is typically preferred in practice for numerical robustness (it

avoids explicit X^{-1}). Both formulations are connected by $H = X^{-1}M + \delta I$, ensuring consistency between theory and implementation.

Algorithm 2: Feasible-Start Primal–Dual IPM (Mehrotra Predictor–Corrector)

Input: $Q \succeq 0$, q, E, tolerances ϵ_{feas}, ϵ_{comp}, maximum iterations *max_iter*

Output: Approximate optimizer (x, y, z)

Step 1: Initialization.;

Construct strictly feasible (x, y, z) as in Algorithm 1;;

$\mu \leftarrow (x^\top z)/n$.;

Step 2: Iterative loop.;

for $k = 1$ **to** *max_iter* **do**

Step 2.1: Residual computation.;

$r_D \leftarrow Qx + q + E^\top y - z; \quad r_P \leftarrow Ex - \mathbf{1}; \quad r_C \leftarrow XZ\mathbf{1} - \mu\mathbf{1}$.;

Step 2.2: Predictor (affine-scaling) step.;

$$\begin{aligned}(Q + X^{-1}Z + \delta I)\,\Delta x_{\text{aff}} + E^\top \Delta y_{\text{aff}} &= -\,r_D - X^{-1}(XZ\mathbf{1}),\\ E\,\Delta x_{\text{aff}} &= -\,r_P,\\ \Delta z_{\text{aff}} &= -\,X^{-1}(XZ\mathbf{1}) - X^{-1}Z\,\Delta x_{\text{aff}}.\end{aligned}$$

Step 2.3: Predictor step sizes and predicted gap.;

Compute $\alpha_{\text{pri}}^{\text{aff}}$, $\alpha_{\text{dual}}^{\text{aff}}$ by the fraction-to-the-boundary rule;;

$\mu_{\text{aff}} \leftarrow \big((x + \alpha_{\text{pri}}^{\text{aff}}\Delta x_{\text{aff}})^\top (z + \alpha_{\text{dual}}^{\text{aff}}\Delta z_{\text{aff}})\big)/n$.;

Step 2.4: Adaptive centering and corrector RHS.;

$\sigma \leftarrow \min\{0.5, \max\{0, (\mu_{\text{aff}}/\mu)^3\}\}$;

$rC_{\text{corr}} \leftarrow XZ\mathbf{1} + (\Delta x_{\text{aff}} \circ \Delta z_{\text{aff}}) - \sigma\mu\mathbf{1}$.;

Step 2.5: Corrector system and directions.;

$$\begin{aligned}(Q + X^{-1}Z + \delta I)\,\Delta x + E^\top \Delta y &= -\,r_D - X^{-1}rC_{\text{corr}},\\ E\,\Delta x &= -\,r_P,\\ \Delta z &= -\,X^{-1}rC_{\text{corr}} - X^{-1}Z\,\Delta x.\end{aligned}$$

Step 2.6: Step-size selection.;

Compute $\alpha_{\text{pri}}, \alpha_{\text{dual}}$ by the fraction-to-the-boundary rule;;

$\alpha \leftarrow \min(\alpha_{\text{pri}}, \alpha_{\text{dual}})$.;

Step 2.7: Update and stopping criterion.;

$x \leftarrow x + \alpha\,\Delta x; \quad y \leftarrow y + \alpha\,\Delta y; \quad z \leftarrow z + \alpha\,\Delta z; \quad \mu \leftarrow (x^\top z)/n$.;

if $\|r_P\|_\infty \le \epsilon_{\text{feas}}$ ***and*** $\|r_D\|_\infty \le \epsilon_{\text{feas}}$ ***and*** $\mu \le \epsilon_{\text{comp}}$ **then**

break

4 Practical Relevance and Applications

An additional motivation for considering interior-point methods lies in their proven effectiveness in large-scale applications. A notable example is provided by Ferris and Munson (2003), who demonstrated the efficiency of primal–dual interior-point algorithms for the training of *Support Vector Machines (SVMs)*, one of the most widely used tools in machine learning. The underlying optimization problem is a quadratic program with relatively few linear constraints and a Hessian matrix Q that can be interpreted as a low-rank update of a positive semidefinite matrix.

In this setting, interior-point methods have proven highly competitive. By exploiting problem structure and employing advanced numerical techniques (such as sparse factorization or the Sherman–Morrison–Woodbury formula), they achieved performance significantly superior to classical alternatives. Reported results include speedups by factors of several hundred and the ability to solve problems with tens of millions of variables through out-of-core implementations that handle datasets larger than memory.

Such examples strengthen the rationale for investigating the *Primal–Dual Feasible Interior-Point Method* in this Project. If these methods can successfully address massive SVM training problems, it is reasonable to expect that they may also be particularly well suited for our constrained quadratic program on a simplicial domain.