Implicit Primal-Dual Interior-Point Methods for Quadratic Programming

Jon Arrizabalaga[1] and Zachary Manchester[1]

***Abstract*— This paper introduces a new method for solving quadratic programs using primal-dual interior-point methods. Instead of handling complementarity as an explicit equation in the Karush-Kuhn-Tucker (KKT) conditions, we ensure that complementarity is implicitly satisfied by construction. This is achieved by introducing an auxiliary variable and relating it to the duals and slacks via a retraction map. Specifically, we prove that the softplus function has favorable numerical properties compared to the commonly used exponential map. The resulting KKT system is guaranteed to be spectrally bounded, thereby eliminating the most pressing limitation of primal-dual methods: ill-conditioning near the solution. These attributes facilitate the solution of the underlying linear system, either by removing the need to compute factorizations at every iteration, enabling factorization-free approaches like indirect solvers, or allowing the solver to achieve high accuracy in low-precision arithmetic. Consequently, this novel perspective opens new opportunities for interior-point methods, especially for solving large-scale problems to high precision.**

I. Introduction

The field of mathematical optimization has evolved significantly over the past several decades, beginning with the advent of the simplex method for linear programming. While simplex methods provided a foundational approach, the development of interior-point methods revolutionized the field by offering polynomial-time guarantees. Initially focusing on primal-only formulations, the community eventually shifted towards primal-dual interior-point methods, which have since become the main driving force for convex programming [1]. These methods are highly regarded because they require very few iterations to converge and possess well-understood theoretical properties [2], [3].

However, due to their fundamental mathematical nature, primal-dual methods inevitably become ill-conditioned near the optimal solution [4]. There is no way around this inherent ill-conditioning, which dictates that expensive matrix refactorizations are required at each iterate, driving up the computational cost per step. Consequently, this heavy reliance on exact factorizations makes traditional primal-dual methods prohibitive for very large-scale problems [5].

This computational bottleneck explains why, in the era of large-scale data, there has been a massive resurgence of first-order methods, most notably in the form of the Alternating Direction Method of Multipliers (ADMM) [6]. These methods are highly amenable to distributed computing and feature remarkably cheap iterations. Nevertheless, they require a vast number of iterations to converge, making it practically impossible to reach high-accuracy solutions with them.

Given this dichotomy—where interior-point methods converge to high accuracy in a few expensive iterations, and first-order methods require many cheap iterations but yield low accuracy—a very reasonable question arises: *Is it possible to develop a method that achieves the low iteration cost of first-order approaches while maintaining the rapid convergence and high precision of interior-point methods?*

In this manuscript, we answer exactly this question by presenting a novel implicit method for solving primal-dual interior-point systems. In particular, our primary contributions are:

1) The introduction of an auxiliary variable and a corresponding retraction map that reformulates the KKT conditions into a linear system that is strictly well-conditioned and bounded.
2) A theoretical proof demonstrating why this specific retraction map is the optimal and unique map capable of guaranteeing the well-conditioning and boundedness of the underlying KKT system.
3) Preliminary results showcasing how our implicit method succeeds in domains where explicit primal-dual methods have traditionally struggled, specifically enabling the reuse of matrix factorizations for inexact Newton steps and facilitating the use of factorization-free iterative solvers.

II. Quadratic Programming

A standard convex Quadratic Program (QP) is given the following primal-dual pair:

$$\text{Primal:} \quad \min_{x} \quad \tfrac{1}{2}x^\top Qx + q^\top x \quad \text{s.t.} \quad Ax \geq b, \quad Cx = d \tag{1a}$$

$$\text{Dual:} \quad \max_{x,\lambda,\gamma} \quad -\tfrac{1}{2}x^\top Qx + b^\top\lambda + d^\top\gamma \quad \text{s.t.} \quad Qx + q - A^\top\lambda - C^\top\gamma = 0, \quad \lambda \geq 0 \tag{1b}$$

where $x \in \mathbb{R}^n$, $s \in \mathbb{R}^m$, $\lambda \in \mathbb{R}^m$, and $\gamma \in \mathbb{R}^p$ are the primal, slack, and dual variables. The problem data consists of the symmetric positive semi-definite objective matrix $Q \in \mathbb{R}^{n\times n}$ ($Q \succeq 0$), the linear objective vector $q \in \mathbb{R}^n$, the inequality constraint matrix $A \in \mathbb{R}^{m\times n}$ with bounds $b \in \mathbb{R}^m$, and the equality constraint matrix $C \in \mathbb{R}^{p\times n}$ with bounds $d \in \mathbb{R}^p$.

[1]Jon Arrizabalaga and Zachary Manchester are with the Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, USA
jonarri@mit.edu, zacm@mit.edu

The Karush-Kuhn-Tucker (KKT) conditions for optimality are:

$$\text{Stationarity:} \quad Qx + q - A^\top \lambda - C^\top \gamma = 0 \tag{2a}$$

$$\text{Primal Feasibility:} \quad Ax - b - s = 0, \quad s \geq 0 \tag{2b}$$

$$Cx - d = 0$$

$$\text{Dual Feasibility:} \quad \lambda \geq 0 \tag{2c}$$

$$\text{Complementarity:} \quad \lambda \odot s = 0 \tag{2d}$$

where $\odot$ denotes the element-wise product. Solving the QP is equivalent to finding the root of this system of equations subject to the non-negativity constraints on λ and s. For clarity we define the global state vector $z = [x^\top, \lambda^\top, \gamma^\top, s^\top]^\top$ and the residual vector as

$$r(z) = \begin{bmatrix} Qx + q - A^\top \lambda \\ Ax - b - s \\ Cx - d \\ \lambda \odot s \end{bmatrix} \tag{3}$$

III. Primal-Dual Interior-Point Methods

Primal-dual interior point methods relax the complementarity condition as

$$\lambda \odot s = \mu, \quad \text{with } \mu > 0, \tag{4}$$

and thus,

$$\begin{bmatrix} Q & -A^\top & -C^\top & 0 \\ A & 0 & 0 & -I \\ C & 0 & 0 & 0 \\ 0 & S & 0 & \Lambda \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta \lambda \\ \Delta \gamma \\ \Delta s \end{bmatrix} = - \begin{bmatrix} Qx + q - A^\top \lambda - C^\top \gamma \\ Ax - b - s \\ Cx - d \\ \lambda \odot s - \mu \end{bmatrix}, \tag{5}$$

where $S = \text{diag}(s)$ and $\Lambda = \text{diag}(\lambda)$. As mentioned before, this step is Newton solve is accompanied by a line search that ensures positivity and progress over a merit function, and an reduction schedule over μ. After some simplifications this can be simplified into the following form:

$$\underbrace{\begin{bmatrix} Q & -A^\top & -C^\top \\ -A & -\Lambda^{-1} S & 0 \\ -C & 0 & 0 \end{bmatrix}}_{E(\lambda, s)} \begin{bmatrix} \Delta x \\ \Delta \lambda \\ \Delta \gamma \end{bmatrix} = \begin{bmatrix} -r_1 \\ r_2 + \Lambda^{-1} r_4 \\ r_3 \end{bmatrix}, \tag{6}$$

Throughout this paper, we will refer to this standard approach as the *explicit* primal-dual method, because the complementarity condition is handled explicitly. The fundamental limitation of this traditional formulation stems from the behavior of the matrix $E(\lambda, s)$ near the optimal solution. As strict complementarity dictates that $\lambda \odot s \to 0$, the elements of the diagonal matrix $\Lambda^{-1} S$ experience extreme scaling disparities—either blowing up to infinity (when $\lambda_i \to 0$) or vanishing to zero (when $s_i \to 0$). This unbounded behavior drives the condition number of $E(\lambda, s)$ to infinity, rendering the linear system highly ill-conditioned and forcing a reliance on expensive, exact matrix factorizations. To directly address this limitation, the following section introduces our *implicit* method, a fundamentally different approach to handling the complementarity condition.

IV. Complementarity via Retraction Maps

To satisfy the complementarity condition in (4) –inspired by [7], [8]– we take an alternative approach, where instead of treating complementarity as an explicit equation that we solve through Newton's method, we augment our state vector by an additional variable $v \in \mathbb{R}^m$, and thereby, enforce complementarity implicitly, i.e., by construction. To achieve this, we introduce a retraction map

$$b_\beta : \mathbb{R} \to \mathbb{R}_+ \text{ such that } b_\beta(v) \cdot b_\beta(-v) = \beta. \tag{7}$$

With this approach our state vector is $z = [x^\top, \lambda^\top, \gamma^\top, s^\top, v^\top]^\top$ and the residual vector as

$$r_\beta(z) = \begin{bmatrix} Qx + q - A^\top \lambda - C^\top \gamma \\ Ax - b - s \\ Cx - d \\ \lambda - b_\beta(v) \\ s - b_\beta(-v) \end{bmatrix}. \tag{8}$$

Notice that, from the definition in (7), the complementarity condition is $\lambda \odot s = \beta$ guaranteed by construction. When linearizing the constraints $\lambda = b(v)$ and $s = b(-v)$ with respect to v, we must apply the chain rule:

$$\nabla_v \lambda = \nabla_v (b_\beta(v)) = db_\beta(v), \tag{9a}$$

$$\nabla_v s = \nabla_v (b_\beta(-v)) = -db_\beta(-v). \tag{9b}$$

where $db(v)$ denote the derivative of b with respect to its argument v. To adopt the matrix form, we diagonalize them as:

$$B_\beta^+(v) = \text{diag}(db_\beta(v)), \quad B_\beta^-(v) = \text{diag}(db_\beta(-v)).$$

The resulting KKT system is

$$\begin{bmatrix} Q & -A^\top & -C^\top & 0 & 0 \\ A & 0 & 0 & -I & 0 \\ C & 0 & 0 & 0 & 0 \\ 0 & I & 0 & 0 & -B_\beta^+(v) \\ 0 & 0 & 0 & I & B_\beta^-(v) \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta \lambda \\ \Delta \gamma \\ \Delta s \\ \Delta v \end{bmatrix} = - \begin{bmatrix} Qx + q - A^\top \lambda - C^\top \gamma \\ Ax - b - s \\ Cx - d \\ \lambda - b_\beta(v) \\ s - b_\beta(-v) \end{bmatrix}, \tag{10}$$

where $B_\beta(v) = \text{diag}(\nabla_v b_\beta(v))$ is the diagonal matrix of the derivative of the retraction map w.r.t. additional variable v. After some simplifications we can reduce this to

$$\underbrace{\begin{bmatrix} Q - A^\top A & -A^\top W_\beta(v) & -C^\top \\ -A & -B_\beta^-(v) & 0 \\ -C & 0 & 0 \end{bmatrix}}_{I(v)} \begin{bmatrix} \Delta x \\ \Delta v \\ \Delta \gamma \end{bmatrix} = \begin{bmatrix} -r_{\beta,1} + A^\top (r_{\beta,2} - r_{\beta,4} + r_{\beta,5}) \\ r_{\beta,2} + r_{\beta,5} \\ r_{\beta,3} \end{bmatrix}. \tag{11}$$

We denote $I(v)$ as the *implicit matrix* and depends on the –yet to be defined– retraction map b_β in (7) and variable v.

A. Choosing the retraction map

To choose the retraction map, we take a principled approach, where we observe the partially condensed form in (11) and identify desirable properties that the retraction map should hold in order to facilitate the solving of the respective linear system. Once we have defined these properties, we proceed to finding retraction map(s) that hold them. We start by assuming that

$$W_\beta(v) = B_\beta^-(v) + B_\beta^+(v) = I\,, \tag{12}$$

then the implicit matrix in (11) becomes

$$I(v) = \begin{bmatrix} Q - A^\top A & -A^\top & -C^\top \\ -A & -B_\beta^-(v) & 0 \\ -C & 0 & 0 \end{bmatrix}, \tag{13}$$

which is symmetric. Additionally if we assume that the retraction map's derivative $B_\beta^-(v)$ is bounded, $I(v,b)$ is always well-conditioned. This is not the case for the standard explicit version $E(\lambda, s)$ in (6), which near the solution is destined to become ill-conditioned.

Following this reasoning, we extend the original definition of the retraction map in (7) as

Definition 1 (Retraction map)**.** *Let $b_\beta : \mathbb{R} \to \mathbb{R}_+$ be a map that satisfies the following properties*

$$b_\beta(v) \cdot b_\beta(-v) = \beta\,, \tag{14a}$$
$$db_\beta(v) + db_\beta(-v) = 1\,, \tag{14b}$$
$$0 < db_\beta(v) \leq 1 \tag{14c}$$

then, b_β is denoted as a retraction map.

Theorem 1. *Let $b_\beta : \mathbb{R} \to \mathbb{R}_+$ be a retraction map satisfying the properties in Definition 1. Then, for any $\beta > 0$, the unique function that fulfills these conditions is the smoothed ReLU function, given by*

$$b_\beta(v) = \frac{v + \sqrt{v^2 + 4\beta}}{2}\,. \tag{15}$$

Proof. From (14b), we have $db_\beta(v) + db_\beta(-v) = 1$. Integrating this equation with respect to v yields

$$b_\beta(v) - b_\beta(-v) = v + C\,, \tag{16}$$

where C is an integration constant. Evaluating (16) at $v = 0$ yields $b_\beta(0) - b_\beta(0) = C$, implying $C = 0$. Thus, $b_\beta(-v) = b_\beta(v) - v$.

Substituting this result into (14a), $b_\beta(v) \cdot b_\beta(-v) = \beta$, we obtain a quadratic equation in terms of $b_\beta(v)$:

$$b_\beta(v)^2 - v b_\beta(v) - \beta = 0\,. \tag{17}$$

Solving for $b_\beta(v)$ using the quadratic formula yields

$$b_\beta(v) = \frac{v \pm \sqrt{v^2 + 4\beta}}{2}\,. \tag{18}$$

By Definition 1, the retraction map maps to strictly positive real numbers ($b_\beta : \mathbb{R} \to \mathbb{R}_+$). Since $\sqrt{v^2 + 4\beta} > |v|$ for any $\beta > 0$, the negative root would yield $b_\beta(v) \leq 0$ for $v \leq 0$. Thus, we must select the positive root.

Finally, we verify that this unique solution satisfies (14c). The derivative is

$$db_\beta(v) = \frac{1}{2}\left(1 + \frac{v}{\sqrt{v^2 + 4\beta}}\right). \tag{19}$$

Because the fractional term is strictly bounded as $\frac{v}{\sqrt{v^2+4\beta}} \in (-1, 1)$ for all $v \in \mathbb{R}$ and $\beta > 0$, it follows directly that $0 < db_\beta(v) < 1$. This satisfies the bounding condition and concludes the proof. □

V. Numerical Experiments

To demonstrate the advantages of our implicit method, we focus on two scenarios where primal-dual interior-point methods excel but explicit formulations struggle. Note that our implementations are prototypical, aiming solely to validate the conceptual benefits of our implicit approach against the standard explicit baseline.

A. Reusing factorization via inexact Newton

Because the retraction map defined in 1 ensures that the KKT system (11) evolves slowly and remains well-conditioned throughout the solving process, we can practically reuse factorizations across multiple iterations. Assuming J^* is the KKT matrix frozen after a previous iterate, we compute a step ensuring sufficient progress by checking:

$$||(J_k - J^*)\Delta z_k|| \leq \eta_k ||r_k||\,. \tag{20}$$

where η_k is the forcing term [9]. This condition is trivial to check for the implicit matrix in (13), as the subtraction occurs strictly between the diagonal elements of the lower-right block. We leverage this to decide whether to compute a new factorization or safely proceed with the frozen one. The results evaluating this strategy are shown in Fig. 1.

B. Factorization-free iterative methods

Iterative Krylov subspace methods enable factorization-free solvers, allowing interior-point methods to scale to massive [10], [11]. However, their notorious sensitivity to ill-conditioning typically demands expensive preconditioners. Our well-conditioned implicit formulation naturally overcomes this bottleneck. By applying MINRES to the explicit system in (6) and our implicit system in (11), Fig. 2 validates the appropriateness and superiority of our approach.

VI. Conclusion

We have presented a novel approach for primal-dual interior-point methods in quadratic programming, where complementarity is guaranteed by construction, i.e., implicitly. The underlying linear system is well-conditioned and bounded, thereby overcoming the ill-conditioning curse of standard explicit methods. This opens new opportunities for solving the underlying linear system, either by reusing factorizations or via factorization-free methods. For these reasons, we believe this implicit representation might be the enabler to bring the benefits of interior-point methods into large-scale optimization problems.

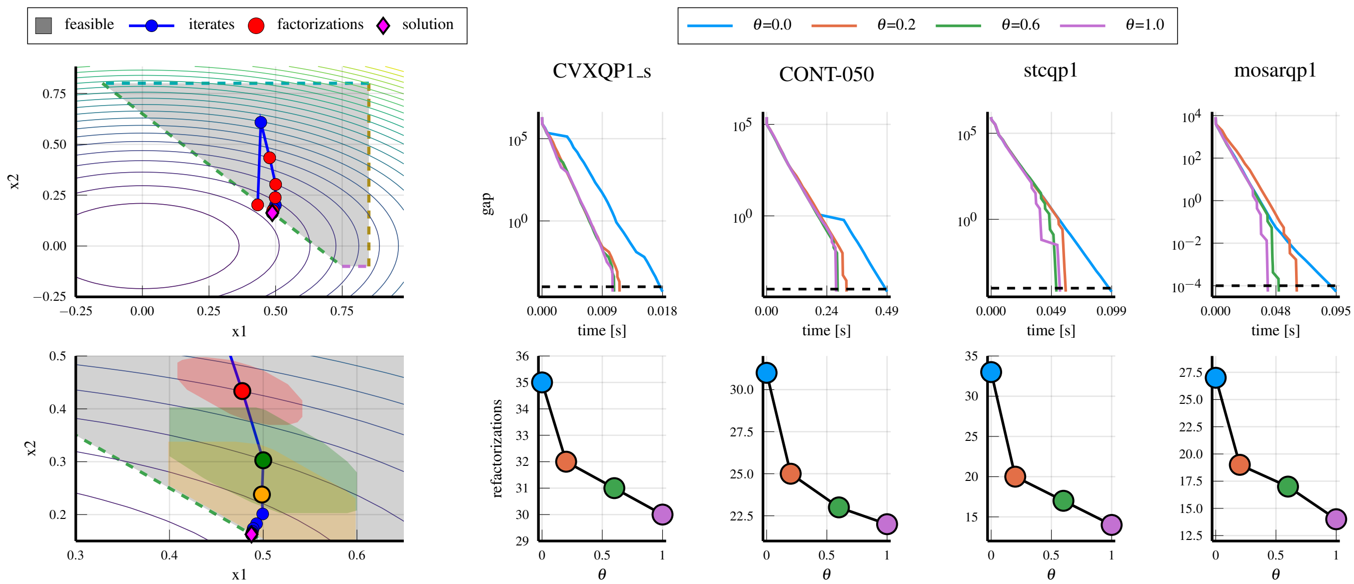

Fig. 1. Evaluation of inexact Newton steps on QP problems from the Maros-Meszaros dataset [12], reducing required factorizations by freezing the KKT matrix. **Column 1:** Synthetic problem iterates (top) and a zoom-in on the final steps (bottom). Red markers highlight explicitly computed factorizations (only 4 of 18 iterations), while the red, green, and yellow regions indicate the validity of the frozen factorizations at those steps. **Columns 2–4:** Performance on the Maros-Meszaros dataset [12]. **Top row:** Duality gap reduction over time for different forcing factors θ. **Bottom row:** Total number of matrix factorizations versus θ. Higher forcing factors enable more aggressive matrix freezing, significantly accelerating overall convergence.

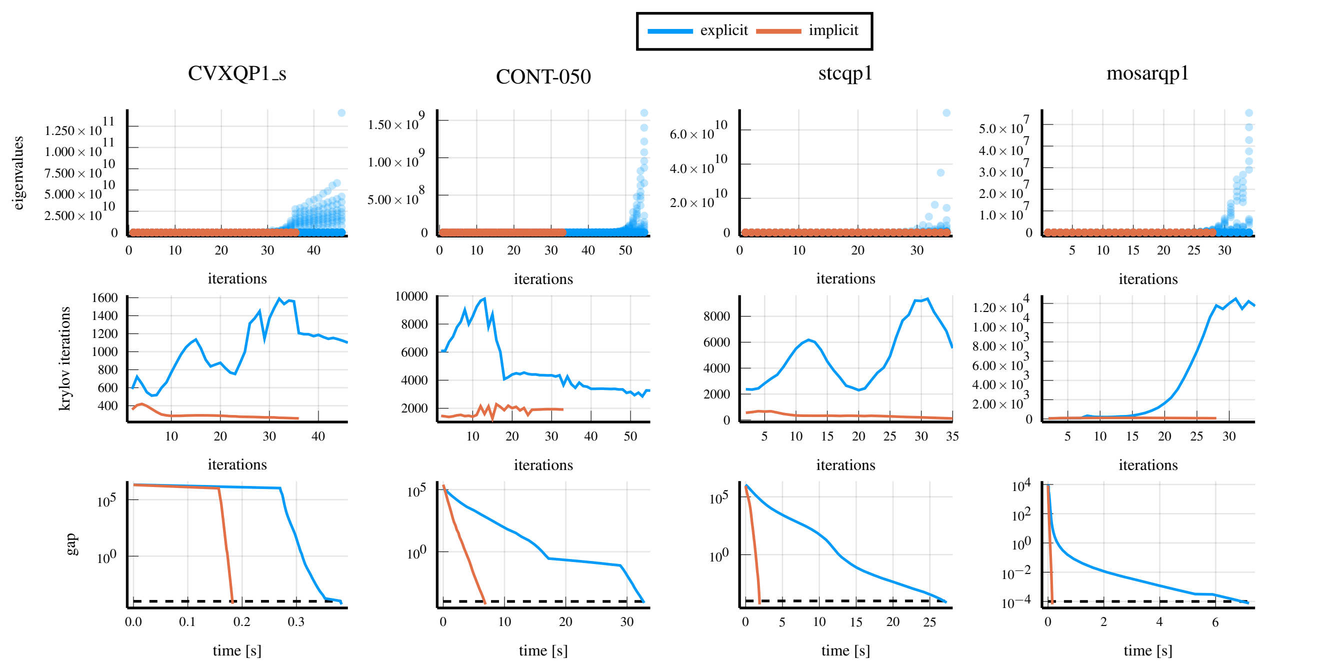

Fig. 2. Evaluation of factorization-free linear system solvers on QP problems from the Maros-Meszaros dataset [12] Factorization-free approaches are crucial for scaling to massive problems where standard matrix factorization becomes computationally prohibitive. Here, we apply the MINRES algorithm to the explicit system in (6) (blue) and our implicit system in (11) (red), utilizing a simple diagonal-Jacobi preconditioner. **Top row:** The singular value spectrum across interior-point iterations. As expected by construction, our implicit formulation remains strictly well-conditioned (singular values remain bounded), whereas the explicit system becomes highly ill-conditioned (singular values explode near the solution). **Middle row:** The number of Krylov (MINRES) iterations required at each interior-point step. Directly following from the condition numbers shown in the top row, the explicit version requires an exploding number of iterations to converge, while our implicit formulation maintains a roughly constant iteration count. **Bottom row:** Duality gap reduction with respect to total solve time. Because the implicit method requires significantly fewer Krylov iterations at each step, its overall solve time is consistently lower.

REFERENCES

[1] J. Gondzio, "Interior point methods 25 years later," *European Journal of Operational Research*, vol. 218, no. 3, pp. 587–601, 2012.

[2] S. J. Wright, *Primal-dual interior-point methods*. SIAM, 1997.

[3] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[4] M. H. Wright, "Ill-conditioning and computational error in interior methods for nonlinear programming," *SIAM Journal on Optimization*, vol. 9, no. 1, pp. 84–111, 1998.

[5] J. Nocedal and S. J. Wright, *Numerical optimization*. Springer, 2006.

[6] P. Neal, C. Eric, P. Borja, and E. Jonathan, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine learning*, vol. 3, no. 1, pp. 1–122, 2011.

[7] F. Permenter, "Log-domain interior-point methods for convex quadratic programming," *Optimization Letters*, vol. 17, no. 7, pp. 1613–1631, 2023.

[8] F. Permenter, "A geodesic interior-point method for linear optimization over symmetric cones," *SIAM Journal on Optimization*, vol. 33, no. 2, pp. 1006–1034, 2023.

[9] R. S. Dembo, S. C. Eisenstat, and T. Steihaug, "Inexact newton methods," *SIAM Journal on Numerical analysis*, vol. 19, no. 2, pp. 400–408, 1982.

[10] H. A. Van der Vorst, *Iterative Krylov methods for large linear systems*. Cambridge University Press, 2003.

[11] F. Liu, M. I. Andersson, A. Fredriksson, and S. Markidis, "Preconditioned cg and gmres for interior point methods with applications in radiation therapy," *Journal of Computational Science*, vol. 90, p. 102 652, 2025.

[12] I. Maros and C. Mészáros, "A repository of convex quadratic programming problems," *Optimization methods and software*, vol. 11, no. 1-4, pp. 671–681, 1999.