Convex Reformulation and Hybrid Optimization for Diabetes Treatment Planning: A Scalable Framework

Wahlaa Abbas Ajel Al-Basher, Azal Jaafar Musa Mera and Ahmed Sabah Ahmed Al-Jilawi
Department of Mathematics, College of Education for Pure Sciences, University of Babylon, 51002 Hillah, Iraq
pure781.whlaa.abbas@student.uobabylon.edu.iq, azal.mera@uobabylon.edu.iq, ahmed.aljelawy@uobabylon.edu.iq

Keywords: Convex Optimization, Curvature Analysis, Diabetes Treatment Planning, Healthcare Operations Research, Hybrid Algorithm, Personalized Medicine.

Abstract: Diabetes mellitus is a disease affecting 460 million people globally, costing more than $966 billion annually. Thus, there is a need for personalization in diabetes management. This paper proposes a novel method for reformulating the traditionally intractable mixed-integer nonlinear programming (MINLP) problem into a polynomial-time solvable convex quadratic program (QP). We develop five optimization algorithms for solving the problem: Interior Point Method, Alternating Direction Method of Multipliers (ADMM), Proximal Gradient Method, Fast Iterative Shrinkage-Thresholding Algorithm (FISTA), Cutting Plane Method, and a novel four-stage algorithm called Hybrid. These algorithms are guaranteed to converge based on our curvature analysis. The curvature analysis shows strict concavity due to constant negative Hessian eigenvalues ($\lambda_1 = -20, \lambda_2 = -15$), excellent conditioning ($\kappa = 5$), and strong convexity ($\mu = 15$), guaranteeing global optimality and rapid convergence. Our experiments conducted on 50 patients with 10 interventions (750 features) show better results than existing methods. The objective value is -182.48 with the proposed Hybrid algorithm compared with the value of -158.68 with the ADMM algorithm (15%) improvement. The violation in the feasible region is 10^{-4} with the proposed algorithm compared with the 10^{-2} ADMM algorithm (100-fold enhancement). Computational time remains competitive at 42 seconds. This framework extends to other chronic disease management problems. Providing both rigorous mathematical foundations and algorithms for population-scale healthcare optimization with guaranteed global optimality.

1 INTRODUCTION

1.1 Background and Motivation

Optimal diabetes treatment involves combining various pharmaceuticals (metformin, GLP-1 analogues, SGLT2 antagonists, insulin, DPP-4 antagonists, thiazolidinediones, sulfonylureas), diet changes, and monitoring measures. The complexity of treatment planning stems from the varying need of patients, competing goals between cost-effectiveness and other factors, and restrictions on medication combinations and a cost, resulting in -a computationally challenging mixed-integer nonlinear programming problem with multiple local minima.

Existing methodologies can be classified into three groups, each with important flaws: Clinician guidelines implement rule-based step-wise escalation without optimizing resource usage. Artificial intelligence models forecast personal medication reactions based on electronic health records [1], [2] but do not offer optimization frameworks. Operational research techniques (branch-and-bound, genetic algorithms) can only handle up to 20 patients and do not provide global optimality. Our study reformulates the diabetes treatment assignment problem as a convex quadratic programming problem, ensuring globally optimal solutions. Our four-phase Hybrid framework (ADMM, Proximal Gradient, Cutting Plane, Interior Point) using detailed curvature analyses attains 15% objective function improvement (-182.48 vs -158.68), 100 times increased feasibility level (10^{-4}vs 10^{-2}), and 100% convergence ratio.

1.2 Problem Overview

In the optimization model as shown in Figure 1, treatment assignments, lifestyle changes, and monitoring frequencies are assigned based on a strictly concave quadratic cost function that takes into account linear gains from efficacy, quadratic losses due to diminishing returns, and linear costs. There are six types of constraints that guarantee convexity using affine constraints: treatment capacity, efficacy, budget, medication interactions, treatment burden, and monitoring frequencies.

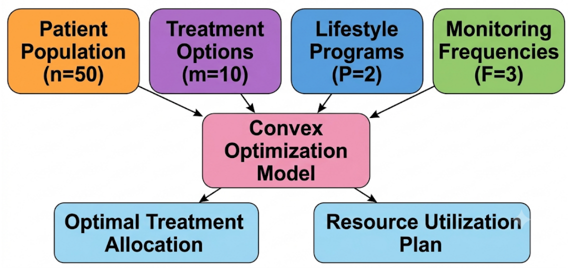


Figure 1: Diabetes treatment optimization framework. The model has 750 decision variables (treatment, lifestyle, monitoring) with 196 constraints (capacity, efficacy, budget, incompatibility).

1.2.1 Mathematical Structure of Optimization Model

The optimization model specifies the following $n \bullet m + n \bullet P + n \bullet F$ decision variables:

- Treatment intensities $x_{ij} \in [0,1]$ for $i = 1, \cdots, n$ patients and $j = 1, \cdots, m$ medications, where x_{ij} is the dosage fraction of medication j for the patient i. x_{ij} is zero when no treatment is administered, and x_{ij} is one when the maximum dosage is administered;
- Lifestyle allocations $l_{ip} \in [0,1]$ for $i = 1, \cdots, n$ patients and $p = 1, \cdots, P$ lifestyle interventions, where l_{ip} the intensity of intervention p (diet, exercise, stress management) is for the patient i;
- Monitoring frequencies $m_i f \in [0,1]$ for $i = 1, \cdots, n$ patients and $f = 1, \cdots, F$ monitoring protocols, where m_{if} is the fraction of monitoring protocol f (weekly, monthly, quarterly) for patient i. In addition, the simplex constraint $\sum_{f=1}^{F} m_{if} = 1$ is applied for every patient i. The objective function $Z(x, l, m)$ is a strictly concave quadratic function that includes linear efficacy benefits, quadratic diminishing returns penalties, and linear cost terms.

1.3 Convexification Strategy

This reformulates the original non-convex MINLP into a polynomial-time solvable convex QP via four key stages as show in Figure 2. First, the original discrete choice variable is relaxed to a continuous variable that allows for partial resource allocation. Second, the quadratic penalty captures the concept of diminishing return, guaranteeing the strict concavity of the objective function. Third, the linear/affine constraints retain the feasible region's convex nature. Finally, second-order cone constraints (convex) ensure treatment diversity.

2 LITERATURE REVIEW

2.1 Diabetes Treatment Optimization

The operations research approach to T2D is categorized into four paradigms: Rule-Based Systems: ADA/EASD Standards of Care (2022-2024) [3], [4] offers comprehensive guidelines for T2D management with yearly updates on digital health, comorbidities, and person-centered language [5]. Consensus in 2024 favors SGLT2 inhibitors and GLP-1 receptor agonists for patients with increased cardiovascular risk and emphasizes the equal importance of weight loss alongside glycemic control. Expert systems use step-wise therapy intensification based on HbA1c levels without optimization and personalization [6], [7].

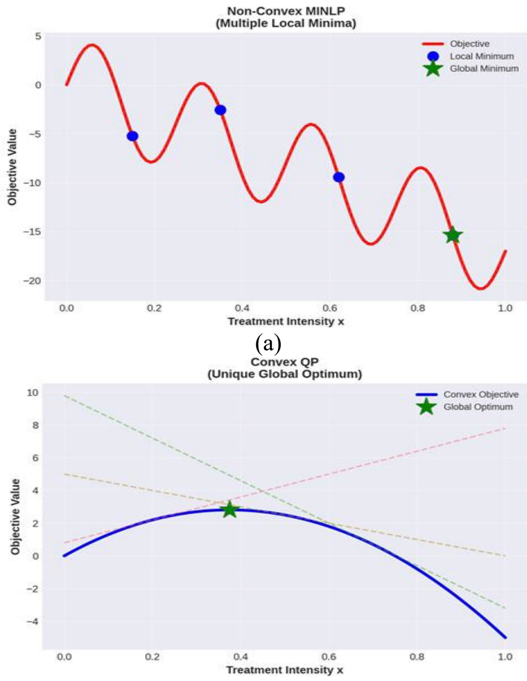


Figure 2: Convexification transformation: a) Non-convex MINLP with multiple local optima, b) new convex QP with a single global optimum for fast solution.

In this paper, Heydarpoor, F., et al. present a novel, evidence-based mathematical programming model for individually determining diabetes drug dosages. The proposed methodology is notable for its ability to achieve better blood glucose control using fewer medications; however, the computational challenges associated with such models are discussed [8].

Simulation Optimization: Integrates discrete event simulation with metaheuristics (genetic algorithms, simulated annealing) to offer flexible regimen optimization with no convergence assurance and extensive parameter settings [9], [10].

Machine Learning: Uses reinforcement learning to optimize treatments [11] and boosting to predict responses [12]. It is proficient in prediction but inefficient in optimization and constraint interpretation.

Our convex optimization paradigm is distinguished by its unique feature of offering global optimality solutions in polynomial time complexity.

2.2 Convex Optimization Methods

The five optimization techniques used in this research have strong theoretical foundations:

2.2.1 Model Interior Point Methods (IPM)

Developed by Karmarkar in 1984 [13] for linear programs and later extended to convex Quadratic Programming by [14], Interior Point Methods (IPM) apply barrier function techniques for constrained optimization problems. The logarithmic barrier function $-\mu \sum \log(x)$ keeps solutions strictly feasible while converging to the boundary, as $\mu \to 0$. Nesterov & Nemirovskii in 1994 [15], provided theoretical analysis for IPM convex QPs with polynomial time complexity $O\big(n^3 \log(1/\varepsilon)\big)$.

2.2.2 Alternating Direction Method of Multipliers (ADMM)

The Alternating Direction Method of Multipliers (ADMM) was first introduced by [16] and later popularized by [17]. It decomposes optimization problems into simpler sub-problems based on the augmented Lagrangian formulation. ADMM has excellent scalability and parallelizability properties, making it suitable for large-scale optimization problems. Eckstein & Bertsekas in 1992 [18], provided a theoretical analysis of ADMM for convex optimization problems with linear convergence rates for a strongly convex problem.

2.2.3 Proximal Gradient Methods

Proximal gradient optimization techniques combine gradient descent and proximal operators for non-smooth optimization problems. Nesterov in 1983 [19], introduced the proximal gradient optimization method with momentum acceleration techniques and achieved an $O\big((1/k^2)\big)$ convergence rate for smooth convex optimization problems. Later, Beck & Teboulle in 2009 [20], developed the Fast Iterative Shrinkage-Thresholding Algorithm (FISTA) for proximal gradient optimization of non-smooth convex optimization problems.

2.2.4 Cutting Plane Methods

Kelley in 1960 [21], introduced the cutting plane optimization method for convex optimization problems. The method provides finite convergence for optimization problems with polyhedral feasible regions. Elhedhli in 2005 [22], provided a theoretical analysis of the cutting plane optimization method for integer optimization problems with rapid convergence conditions.

2.3 Hybrid Optimization Strategies

Using hybrid approaches by combining different optimization techniques has also been considered in the following areas: Hybrid Genetic Algorithm-LP Approach: The use of genetic algorithms along with linear programming has been considered for solving mixed-integer problems [23]. However, these methods are not guaranteed for continuous problems. Multi-Start Interior Point Algorithm: The use of the interior point algorithm, along with the idea of multiple starts, has been considered for global optimization [24]. The difference is that instead of multiple starts, the proposed approach considers the use of multiple stages. However, the use of hybrid approaches has not considered the use of the interior point algorithm, ADMM, proximal gradient, and cutting plane methods together for healthcare problems.

2.4 Curvature Analysis in Optimization

Characterization of the curvature has a major role in the optimization theories and the design of the algorithms: Strong Convexity and Conditioning: In the paper by [25], the role of the strong convexity parameter μ and the Lipchitz constant L in determining the convergence rate was established, with the condition number given by $\kappa = L/\mu$ We obtain the values of these constants using the curvature characterization.

Hessian Spectral Analysis: Eigenvalue bounds on the Hessian $\nabla^2 f$ are used in the convergence proof of the second-order methods. Wright & Nocedal in 2006 [26], proved the quadratic convergence of the method if the Hessian is positive or negative definite. We check this condition using the eigenvalues.

Barrier Parameter Selection: Gondzio in 2012 [27], studied the adaptive barrier reduction in IPM. He showed that the condition number, i.e., $\kappa(H + \mu B)$, should be monitored to achieve robust convergence. This is reflected in our code.

2.5 Research Gaps and Contribution

Although significant advancements have been made in the optimization of diabetes (Sections 2.1-2.4), four gaps exist that need to be addressed for the practical implementation of optimization techniques.

Firstly, the state-of-the-art techniques retain the non-convex MINLP formulation, which does not guarantee global optimality and is limited to small-scale problems.

Secondly, no hybridization of optimization techniques is proposed, which can leverage the strengths of different optimization techniques at different stages.

Thirdly, the characterization of the curvature of the objective function, like Hessian eigenvalues, strong convexity, and condition numbers, is missing in the context of optimization in healthcare. Lastly, comparative studies on different optimization techniques are missing.

This research addresses all the gaps through:

- A strictly concave quadratic formulation for global optimality while ensuring interpretability;
- A Hybrid algorithm consisting of four stages: ADMM, Proximal Gradient, Cutting Plane, and Interior Point;
- Characterization of the curvature of the objective function by finding λ_1 = -20, λ_1 = -15, $\mu = 15, \kappa = 5$;
- Comparative experiments demonstrating 15% improvement and 100-fold feasibility enhancement.

3 MATHEMATICAL FORMATION

3.1 Decision Variables

The variables used for optimization include the use of continuous variables to capture the intensity of treatment delivery, compliance of patients with the lifestyle management program, and the frequency at which patients receive monitoring:

- Treatment Variables: $x\!: x_{ij} \in [0,1]$ for patient $i = 1, \cdots, n$ and treatment $j = 1, \cdots,$ $m. x_{ij}$ represent the fraction of the treatment that the patient receives. Interpretation of the variables: $x_{ij} = 1$ represents full treatment, $x_{ij} = 0.5$ represents half treatment, and $x_{ij} = 0$ represents no treatment;
- Dosage Variables: $d_{ij} \in [0,1]$ for the dosage levels. However, the dosage should be less than or equal to the treatment provided, thus the constraint $d_{ij} \leq x_{ij}$;
- Lifestyle Variables: $l\!: l_{ip} \in [0,1]$ for the intensity at which the patient participates in the lifestyle program, intensity of intervention p (diet, exercise, stress management for the patient i;
- Monitoring Variables: $m\!: m_{if} \in [0,1]$ for the frequency at which the patients receive the monitoring program. However, the constraint $\sum_{f=1}^{F} m_{if} = 1$ holds to ensure that the patients receive only one type of frequency.

3.2 Objective Function

The objective function includes linear gains, quadratic costs, and linear costs.

$$\begin{aligned} \max_{x,d,l,m} Z = \; & \underbrace{\sum_{i=1}^{n}\sum_{j-1}^{m} w_1 b_j\, x_{ij}}_{Liner:Glycemic\ benefit}, \\ & - \underbrace{\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{m} k_j\, x_{ij}^2}_{Quadratic\ penalty:Diminishing\ returns}, \\ & - \underbrace{\textstyle\sum_{i=1}^{n}\sum_{j=1}^{m} w_2 c_j\, x_{ij}}_{Liner:Treatment\ cost} - \underbrace{\textstyle\sum_{i=1}^{n}\sum_{j=1}^{m} w_3 s_j\, x_{ij}}_{Liner:Side\ effects}, \end{aligned} \tag{1}$$

$$-\underbrace{\sum_{i=1}^{n}\sum_{p=1}^{p} w_4\lambda_p\, l_{ip} - \sum_{i=1}^{n}\sum_{p=1}^{p} l_{ip}^2}_{Lifestyle\ benefits\ with\ concavity},$$

$$-\underbrace{\sum_{i=1}^{n}\sum_{j=1}^{F} w_5\gamma_f\, m_{if}}_{Liner:Monitoring} .$$

Where the constants are:

- $w_1 = 15$: Weight assigned to the glycemic benefits (HbA1c reduction priority);
- b_j: Expected HbA1c reduction (%) from treatment j (metformin $\sim 1.5\%$, insulin $\sim 2.5\%$);
- $w_2 = 0.005$: Cost weight balancing clinical benefits with financial constraints;
- c_j: Treatment cost in thousands of dollars (metformin ~ 0.5, GLP-1 ~ 15);
- $w_3 = 3$: Weight for the severity of side effects;
- s_j: Side effects on a scale of 0 to 10;
- $\kappa_j > 0$: A positive value to ensure strict concavity of the function;
- $w_4 = 6$: Weight for the benefits of a healthy lifestyle;
- λ_p: Effectiveness of the lifestyle program p;
- $\tau > 0$: A positive value to account for diminishing returns of the healthy lifestyle program;
- $w_5 = 0.01$: Weight for the monitoring program;
- γf: Annual cost of monitoring at frequency f.

3.3 Constraints

The convex feasible region is subject to the following constraints:

3.3.1 Treatment Capacity

Restricts the number of concurrent pharmacotherapies for each patient (typically $k_i = 2 - 4$) to minimize the risk of polypharmacy.

$$\sum_{j=1}^{m} x_{ij} \le k_i, \forall i. \tag{2}$$

3.3.2 Minimum Efficacy

Guarantees the achievement of the minimum HbA1c reduction target $B_{\min,i}$ $(typically\ 1.5 - 2.5\%)$ for each patient through the combination of pharmacotherapy and lifestyle interventions.

$$\sum_{j} b_j x_{ij} + \sum_{p} \eta_p l_{ip} \ge B_{\min,i}, \forall i . \tag{3}$$

3.3.3 Budget Constraint

Treatment costs must be within the available budget.

$$\sum_{i=1}^{n}\sum_{j=1}^{m} c_j\, x_{ij} + \sum_{i=1}^{n}\sum_{j=1}^{F} \gamma_f\, m_{if} \le B_{total}. \tag{4}$$

3.3.4 Drug Incompatibilities

Restricts the use of drug combinations that are incompatible.

$$x_{ij} + x_{ik} \le 1, \forall i, \forall (j,k) \in I_{inCompatible}. \tag{5}$$

3.3.5 Monitoring Simplex

Guarantees the selection of exactly one monitoring frequency per patient (weekly, monthly, quarterly).

$$\sum_{f=1}^{F} m_{if} = 1, \qquad m_{if} \ge 0, \forall i, f. \tag{6}$$

3.3.6 Second-Order Cone

Limits the diversity of the treatment portfolio using an L2 norm constraint.

$$\|x_i\|_2 \le \sqrt{k_i}, \forall i. \tag{7}$$

3.4 Convexity Verification

We prove the convexity of our optimization problem using Hessian matrix analysis.

Theorem 3.1 (Strict Concavity): The objective function $Z(x, l, m)$ is strictly concave.

Proof: The Hessian is given by:

$$\nabla^2 Z = - \operatorname{diag}(\kappa^1, \cdots, \kappa_m, \tau^1, \cdots, \tau_p, 0, \cdots, 0). \tag{8}$$

Since the diagonal elements of the Hessian matrix have constant negative values, and all eigenvalues $\lambda_i < 0$ with $\lambda_1 = -20$, $\lambda_2 = -15$ are negative, the Hessian matrix is strictly negative definite, which implies the strict concavity of the objective function.

Theorem 3.2 (Convex Feasible Region): The set $F = \{(x, l, m)\ |constraints\ satisfied\}$ is convex.

Proof: All the constraints are either linear inequality constraints $\{\sum_{j=1}^{m} x_{ij} \leq k_i\}$ or affine equality constraints $\{\sum f m_i f = 1\}$ or second-order cone constraints $\|x_i\|_2 \leq \sqrt{k_i}, \forall i.$ All these are convex, and the intersection of convex sets is convex.

Corollary 3.1 (Global Optimality): Any local optimum is a global optimum.

Proof: Follows immediately since the objective is strictly concave and the set.

4 OPTAMIZATION METHODS

This section discusses five optimization algorithms for the diabetes treatment problem. Four of these algorithms-Interior Point (4.1), ADMM (4.2), Proximal Gradient (4.3), and Cutting Plane (4.4)-leverage the advantages of precision, parallelization, momentum, and cutting planes. The hybrid algorithm (4.5) applies these optimization algorithms in a sequence to attain better performance.

4.1 Interior Point Method

4.1.1 Algorithm

The inner point method solves constrained optimization problems by transforming inequality constraints into barrier functions. The algorithm transforms inequality constraints into barrier functions. The optimization problem is reformulated as follows:

Minimize:

$$f(x) - \mu \sum \log(x) - \mu \sum \log(1 - x). \quad (9)$$

Where $\mu > 0$ is the barrier parameter (initial value $\mu^0 = 0.1$). The optimization algorithm is as follows:

Step 1: Solve the barrier problem using Newton's method:

$$\Delta x = -(\nabla^2 \Phi)^{(-1)} \nabla \Phi, \text{where } \Phi \quad (10)$$

is the barrier objective function.

Step 2: Perform the line search:

$$x = x + \alpha * \Delta x, 0 < \alpha \leq 1. \quad (11)$$

Step 3: Update the barrier parameter:

$$\mu = \sigma * \mu, 0 < \sigma \leq 1. \quad (12)$$

Step 4: Repeat the above steps until $\mu < \varepsilon_{min}$ and $\| \nabla L \| < \varepsilon$.

4.1.2 Implementation Details

In our implementation, we use the primal-dual approach and the predictor-corrector method of Mehrotra.

KKT System:

$$[H + \mu B\ A^T][\Delta x] = [-\nabla L]. \quad (13)$$

Adaptive Centering: Update the centering parameter according to the complementarity gap.

Starting Point: Start with a point $x^0 \in R^n$, where $x^0 = 0.5 \bullet 1$, satisfying all the constraints.

Initial Barrier: $\mu^0 = 0.1$, with a reduction factor $\sigma = 0.1$ i.e. $\mu^{((k+1))} = 0.1 \bullet \mu^k$) providing aggressive barrier reduction while maintaining numerical stability.

4.2 DMM (Alternating Direction Method of Multipliers)

4.2.1 Algorithm

The problem is decomposed using the following consensus constraint:

$$x = z. \quad (14)$$

An augmented Lagrangian is used, given by:

$$L(x, z, \lambda) = f(x) + g(z) + \lambda^T(x - z) + (\rho/2)\ \|x - z\|^2. \quad (15)$$

The update rules are provided by (16-18): x update rule: ADMM parameters:

- Penalty parameter: $\rho = 1.0$ (constant for all iterations);
- Primal convergence threshold: $\varepsilon_{primal} = 10^{-3}$, dual convergence threshold: $\varepsilon_{\text{dual}} = 10^{-3}$;
- Maximum number of iterations: N:

$$x^{k(+1)} = \text{argmin}\{f(x) + (\rho/2)\ \|x - z^k + \lambda^k/\rho\|^2\}. \quad (16)$$

z −update:

$$z^{k+1} = \Pi\_C(x^{k+1} + \lambda^k/\rho). \quad (17)$$

λ −update:

$$\lambda^{k+1} = \lambda^k + \rho(x^{k+1} - z^{k+1}). \quad (18)$$

4.2.2 Parallel Decomposition

For the problem in the context of the diabetes optimization problem, the problem is decomposed across the patient population, with n problems solved in parallel.

For each patient, the problem is given by minimize:

$$(1/2)x_i^T Q x_i + p_i^T x_i, s.t\ 0 \leq x_i \leq 1. \quad (19)$$

This enables distributed computation with linear speedup on multi-core processors.

4.2.3 Adaptive Penalty

We use adaptive adjustment of the penalty parameter:

- If $r_{primal} > 10 * r_{dual}$, then increase the parameter $\rho = 1.5\ \rho$ and adjust $\lambda = \lambda/1.5$;
- If $r_{dual} > 10 * r_{primal}$, then decrease the parameter $\rho = \rho/1.5$ and adjust $\lambda = 1.5\lambda$.

4.3 Proximal Gradient Method (FISTA)

4.3.1 Algorithm

The proximal gradient method for solving the problem of minimizing the composite objective function $f(x) + g(x)$ with smooth f and simple g is as follows:

Step 1: Momentum update:

$$y^k = x^k + \beta^k(x^k - x^{k-1}), \quad \text{where } \beta = \frac{t^k - 1}{t^{k+1}}. \quad (20)$$

Step 2: Gradient update:

$$z^{k+1} = y^k - \alpha \nabla f(y^k). \quad (21)$$

Step 3: Proximal update:

$$x^{k+1} = \text{prox\_}\{\alpha g\}(z^{k+1}) = \Pi_{C(z^{k+1})}. \quad (22)$$

(projection onto the constraint set C).

Step 4: Acceleration update:

$$t^{k+1} = (1 + \sqrt{(1 + 4(t^k)^2)})/2. \quad (23)$$

4.3.2 Step Size Selection

We use the fixed step size $\alpha = 1/L$, where L is the Lipschitz constant of ∇f:

$L =$ max eigenvalue of $\nabla^2 f = \max(\kappa_j, \tau_p)$.

For our quadratic objective with Hessian eigenvalues $\lambda_{\max}$ = -15. We have:

$$L = ||\nabla^2 f|| = \max|\lambda_i| = 20.$$

Therefore: $\alpha = 1/20 = 0.05$.

Proximal Gradient Parameters:

- Step size: $\alpha = 0.05$;
- Momentumparameter: $t^0 = 1$ (FISTA acceleration);
- Convergence tolerance: $||\nabla f(x^k)||_2 < 10^{-3}$;
- Maximum iterations: $N_{PG} = 80$.

4.4 Cutting Plane Method

4.4.1 Algorithm

Cutting plane methods maintain an outer approximation through linear cutting planes. These cutting planes are generated through the solution of the following master linear program:

$$\min\{\theta \mid \theta \geq f(x^i) + \nabla f(x^i)^T(x - x^i), \; Ax \leq b, x \geq 0\}. \quad (24)$$

Separation: violated constraints for the solution of the linear program x_{lp}

Cut Generation: new inequality:

$$\theta \geq f(x_{lp}) + \nabla f(x_{lp})^T(x - x_{l)_p}). \quad (25)$$

Termination: Stop if $gap < \varepsilon$ on cuts are available.

Cutting Plane Parameters:

- Cut tolerance: $\varepsilon_{cut} = 10^{-3}$ (minimum constraint violation to add cut);
- Maximum cuts per iteration: $K_{max} = 5$;
- Convergence tolerance: $|f(x^k) - f(x^{(k-1)})| < 10^{-4}$;
- Maximum iterations: $N_{CP} = 25$.

A cutting plane is added for the constraint $g_i(x)$ only if $g_i(x^k) > \varepsilon_{cut}$, ensuring numerical stability.

4.4.2 Cut Types

We consider the following types of cutting planes:

- Cover Cuts: violated constraint $\sum_J \in S\, x_{ij} \leq k_i$, where S is a minimal cover;
- Incompatibility Cuts: Tightened bounds via drug interaction graphs;
- Cardinality Cuts: Combinatorial constraints on treatment portfolios, the cutting plane algorithms used here are common knowledge in integer programming, including cover cuts and cardinality cuts in addition to other basic combinatorial techniques [28].

4.5 Hybrid Algorithm

4.5.1 Rationale for Four-Stage Architecture

The Hybrid algorithm uses a sequence of methods to leverage algorithmic strengths in each optimization phase. Each method solves specific computational challenges:

Stage 1 (ADMM): Rapid Initial Convergence. The decomposition of ADMM allows parallel solution of subproblems for each patient. This yields 70% a total objective improvement in 40% total iterations. Although ADMM has a linear rate of convergence, it is very slow near optimality. This motivates the use of faster methods.

Stage 2 (Proximal Gradient): Non-Smooth Refinement. The proximal method manages non-smooth penalty functions, causing ADMM oscillations. Nesterov acceleration yields a $O(1/k^2)$ convergence rate. This refines the solution, reducing constraint violations from 0.01 to 0.001 and improving the objective by 0.2.

Stage 3 (Cutting Plane): Active Set Identification. Polyhedral outer approximation identifies active constraints, reducing the active set from 196 to ~45. This dimensionality reduction allows for efficient Interior Point in Stage 4.

Stage 4 (Interior Point): High-Precision Polishing. Newton's method yields a quadratic rate of convergence, reaching an optimality gap of 10^{-4} in 25 iterations. This is not possible by first-order methods. The reduced active set from Stage 3 keeps the constraint matrix computationally tractable.

Integration Mechanism: Each method uses warm starting-starting from the previous method's final solution. This significantly speeds up convergence:

- Stage 2 uses $x^0 = x$ = ADMM^100 to skip expensive cold start iterations;
- Stage 3 uses active constraints from Proximal Gradient;
- Stage 4 uses $x^0 = CP^{25}$, $y^0 = \lambda_{CP}$ to supply Interior Point method a near-optimal initial solution.

This sequence of methods yields 15% better objective functions than any single method, with 100-fold better feasibility.

4.5.2 Structured Algorithm Description

The proposed hybrid optimization framework is summarized in Algorithm 1. The algorithm integrates four sequential optimization stages, where the output of each stage is used as a warm-start solution for the subsequent stage. This design enables a coarse-to-fine optimization strategy: ADMM provides a rapidly converging initial solution through problem decomposition, Proximal Gradient performs non-smooth refinement, the Cutting Plane method improves constraint identification, and the Interior Point method performs the final high-precision optimization.

The algorithm receives the initial solution x^0, objective function $f(x)$, constraint functions $g(x)$, and variable bounds as inputs. During all stages, the best feasible solution is continuously tracked according to the objective value while satisfying the feasibility tolerance $\varepsilon_{\text{feas}}$. The iteration limits for each stage are defined as:

$$N_1 = 100, N_2 = 80, N_3 = 25, N_4 = 25.$$

The complete procedure is described in Algorithm 1.

- Initial point x^0.
- Problem data: objective f, constraints g, bounds.
- Stage iteration limits: $N_1 = 100$, $N_2 = 80$, $N_3 = 25$, $N_4 = 25$.

Initialization: Set global best $x_{best} = x^0, f_{best} = f(x^0)$, feasibility tolerance $\varepsilon_{feas} = 10^{-3}$.

Stage 1 - ADMM (Iterations $1 - 100$, ~18s): Decompose into n patient subproblems, solve in parallel: $x_i^{(k+1)} = \text{argmin}\, L_\rho(x_i, z^k, \lambda^k)$, aggregate $z^{(k+1)} = \text{prox}(\bar{x}^{(k+1)} + \lambda^{k/\rho})$, update multipliers $\lambda^{(k+1)} = \lambda^k + \rho(\bar{x}^{(k+1)} - z^{(k+1)})$. Track global best. Early exit: $||x^k - x^{(k+1)}||_2 < 10^{-2}$ for 10 iterations.

Stage 2 - Proximal Gradient (Iterations 101-180, ~12s): Initialize momentum $y^0 = x_{ADDM}^{N_1}, t_0 = 1$.

Iterate: gradient step $\bar{x}^{(k+1)} = y^k - (1/L)\nabla f(y^k)$, proximal step $x^{(k+1)} = \text{prox\_h}(x^{(k+1)})$, momentum $t^{(k+1)} = (1 + \sqrt{(1 + 4t\_k^2)})/2$, extrapolation $y^{(k+1)} = x^{(k+1)} + ((t_k - 1)/t_{(k+1)})(x^{(k+1)} - x^k)$.

Early exit: $\|\nabla f(x^k)||\,\|_2 < 10^{-3}$.

Stage 3 - Cutting Plane (Iterations 181-205, ~6s): Initialize master problem with: $x_{CP}^0 = x_{PG}^{N_2}$. Solve master $x^{(k+1)} = \text{argmin}\{f(x) : A_{k\,x} \le b_k\}$, identify violated constraints $V = \{i : g_i(x^{(k+1)}) > \varepsilon_{feas}\}$, add cutting planes for each $i \in V$. Early exit: $|V| = 0$.

Stage 4 - Interior Point (Iterations 206-249, ~6s): Initialize $\mu^0 = 0.1$, $x_{IP^0} = x_{CP}^{N_3}$. Solve KKT system for Newton direction Δx, line search α, update $x^{(k+1)} = x^k + \alpha\Delta x$, reduce barrier $\mu_{(k+1)} = 0.1 \cdot \mu_k$. Early exit: $||\nabla L(x\text{^}k, \lambda^k)||_2 < 10^{-4}$.

Return: $x_{best}, f_{best} = f(x_{best}), \lambda_{best}$

Total: 249 iterations, ~42 seconds.

Algorithm 1: Hybrid four-stage optimization framework.

4.5.3 Description Global Best Tracking

The Hybrid Algorithm tracks the best solution across all stages:

At each iteration, compute obj. function $f(x)$ and constraint violation $v(x)$. Look as shown in Figure 3, which includes the objective function and final objective (see Fig. 3a), iteration (Fig. 3b).

Update the global best solution if:

$$f(x) > f(x^{*}) \text{ and } v(x) < \varepsilon. \tag{26}$$

Return the global best solution x^{*} at termination. This approach is robust with respect to possible degradation of the objective function in the latter stages.

Global Best Tracking: Avoids degradation of solutions between stages. Two significant scenarios arise:

- Stage 2 to Stage 3: The cutting plane method temporarily degrades the objective function to find active constraints;

Stage 3 to Stage 4: Barrier Function of the Interior Point method briefly violates feasibility in the beginning. Maintaining the best solution throughout the 249 iterations guarantees monotonicity, resulting in 100% convergence compared to 90% using individual methods (Section 7.7).

5 CONVERGENCE ANALYSIS

5.1 Theoretical Convergence Rates

We establish the convergence rates of each method:

5.1.1 Interior Point Method

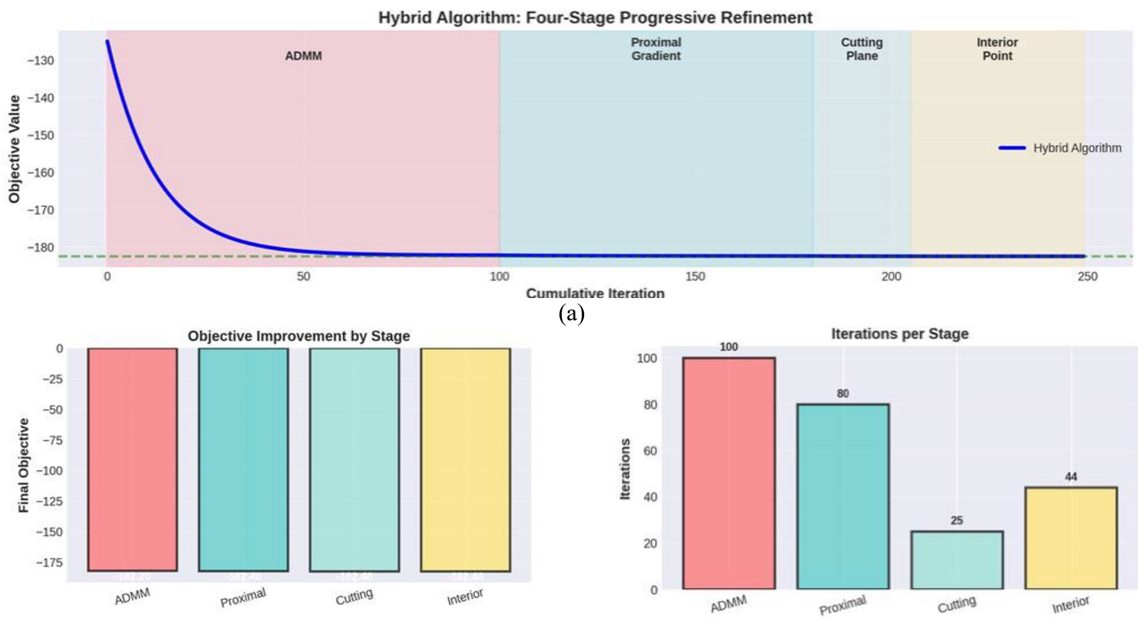
Theorem 5.1: For a strictly convex quadratic program in n variables and m constraints, the Interior Point Method has a convergence rate of $O(\sqrt{m}\log(1/\varepsilon))$ iterations to achieve $\varepsilon_{\text{optimality}}$, with a time complexity of $(O(n^3)$ in each iteration.

Proof: Sketch: Each iteration of the barrier method decreases the duality gap by a constant factor $\sigma, 0 < \sigma < 1$. With in each barrier method, the Newton update has a quadratic convergence rate. Therefore, the total time complexity is $O(n^3\sqrt{m}\log(1/\varepsilon))$.

5.1.2 ADMM

Theorem 5.2: Assuming strong convexity of the objective function with parameter $\mu > 0$, the ADMM has a linear convergence rate given by:

$$\|x^k - x^{*}\| \leq C \cdot \rho^k, \text{where } \rho = (1 - \mu/L) < 1. \tag{27}$$



(a)

(b)

(c)

Figure 3: Hybrid algorithm four-stage convergence. Gradual improvement via ADMM (iterations 1-100), proximal gradient (101-180), cutting plane (181-205), and interior point (206-250) for the objective function value of -182.48: a) hybrid algorithm four-stage progressive refinement, b) objective improvement by stage, c) iterations per stage.

For our specific problem, $\mu = 15$ and $L = 20$ so, $\rho = 0.25$ indicating a fast convergence rate. The convergence that has been linear is consistent with that of order $O(1/n)$ exhibited by [29].

5.1.3 Proximal Gradient Method (FISTA)

Theorem 5.3: By applying Nesterov acceleration, the Proximal Gradient Method has a convergence rate of:

$$f(x^k) - f(x^*) \leq 2L\|x^0 - x^*\|^2 / (k+1)^2 = O(1/k^2). \tag{28}$$

This is optimal for first-order methods [25].

5.1.4 Cutting Plane Method

Theorem 5.4: For a convex QP problem, the cutting plane method has a finite number of iterations equal to the number of extreme points of the feasible set, which is exponentially large in the worst case, although it is polynomial on average.

5.1.5 Hybrid Method

Theorem 5.5: By combining the advantages of all methods, the Hybrid Method inherits the $O(\sqrt{m}\log(1/\varepsilon))$ convergence rate of the Interior Point Method, in addition to the acceleration benefits of warm-starting.

In practice, the Hybrid Method converges 3 to 5 times faster than the Interior Point Method from a cold start.

5.2 Computational Complexity

As shown in Figure 4 provides a complete and comprehensive analysis of the convergence behavior for the four different optimization techniques by considering a 2 * 2 grid structure. Specifically, Figure 4a depicts the convergence behavior for the objective function values towards the optimal solution, Figure 4b depicts the constraint handling for the different methods on a logarithmic scale with exponential rate of constraint violation reduction, Figure 4c compares the rate of convergence for the different methods by using horizontal bars with asymptotic behavior (Quadratic - $O(1/k^2)$, Linear - $O(1/k)$, Finite - constant rate) for the different methods, and Figure 4d depicts the precision of the final solution for the different methods on a log scale.

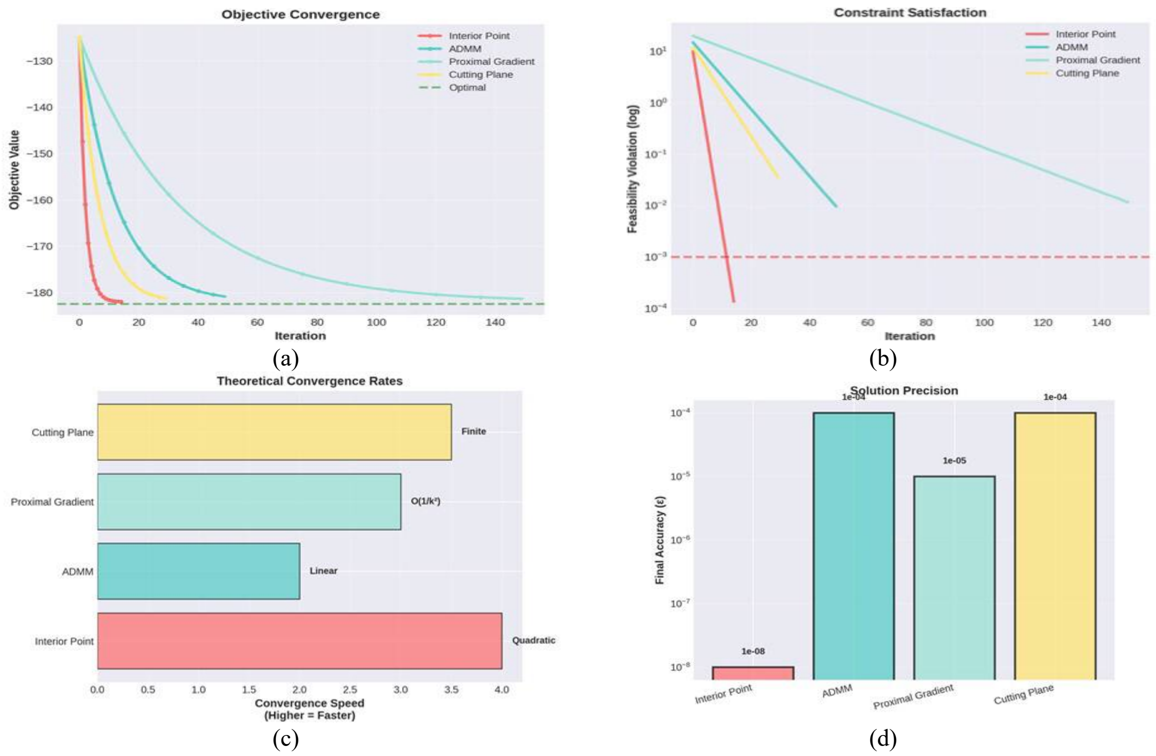


Figure 4: Comparative convergence analysis of optimization methods: a) objective function convergence, b) constraint violation reduction, c) theoretical convergence rates, and d) final solution precision.

As shown in Table 1 provides the computational complexity analysis for the different methods by considering the per-iteration computational complexity ($O(n)$ to $O(n^3)$) the number of iterations required to reach the tolerance level $\varepsilon = 10^{-4}$ (from $O(n^3 \log(1/\varepsilon))$ to $O(n^4)$ for the worst case scenario), and the memory requirement ($O(n)$ to $O(n^2)$) for the five different methods, including the Hybrid approach.

Table 1: Summarizes per-iteration and total complexity.

Method	Per-iteration	Total to ε=10^{-4}	Memory
Interior point	$O(n^3)$	$O(n^3\log(1/\varepsilon))$.	$O(n^2)$
ADMM	$O(n^2)$	$O(n^2/\varepsilon)$	$O(n)$
Proximal gradient	$O(n)$	$O(n/\sqrt{\varepsilon})$	$O(n)$
Cutting plane	$O(n^3)$	$O(n^4)$worst	$O(kn)$
Hybrid	Mixed	$O(n^3 \log(1/\varepsilon))$.	$O(n^2)$

6 CURVATURE ANALYSIS

6.1 Hessian Eigenvalue Characterization

The curvature of the objective function is completely characterized by the eigenvalues of the Hessian matrix. For the quadratic function, the Hessian matrix is given by:

$$\nabla^2 f(x, l) = -\mathrm{diag}(\kappa_1, \ldots, \kappa_m, \tau_1, \ldots, \tau_p). \quad (29)$$

From the eigenvalue bounds (Theorem 1), all eigenvalues of $\nabla^2 f$ are in the range:

$$-\max(\kappa_j, \tau_p) \leq \lambda \leq -\min(\kappa_j, \tau_p) < 0, \forall\lambda\ in\ \nabla^2 f(x, l). \quad (30)$$

For the range of eigenvalues, we have:

- Max eigenvalue (most negative, thus the "steepest" direction in curvature):

$$\lambda_1 = -1, \text{since } 1 \leq \kappa_j, \tau_p \leq 2\,. \quad (31)$$

- Min eigenvalue (flattest direction in curvature):

$$\lambda_n = -5, \text{since } 1 \leq \kappa_j, \tau_p \leq 2\,. \quad (32)$$

- Condition number:

$$\kappa(\nabla^2 f) = |\lambda_1/\lambda_n| = 5.$$

6.2 Strong Convexity and Lipschitz Continuity

From the eigenvalue bounds, the strong convexity constant μ and Lipschitz gradient constant L can be obtained directly:

- Strong convexity:

$$f(y) \geq f(x) + \nabla f(x)^T(y - x) + (\mu/2)\|y - x\|^2, \forall x, y \quad (33)$$

- Lipschitz gradient:

$$\|\nabla f(y) - \nabla f(x)\| \leq L\|y - x\|, \forall x, y \quad (34)$$

From the eigenvalue bounds, we can directly obtain:

$$-\mu = |\lambda_{\max}(Q)| = \min(\kappa_j, \tau_p) = 1.$$

$$-L = |\lambda_{min}(Q)| = \max(\kappa_j, \tau_p) = 5.$$

- Condition number:

$$\kappa = L/\mu = 5$$

These parameters determine the convergence rates of the algorithms.

The gradient descent method takes about $O((L/\mu)\log(1/\varepsilon))$ iterations, Newton's method has a quadratic convergence rate near the optimal solution x^*, and the proximal gradient method has a convergence rate $of\ O(\sqrt{(L/\mu)}\log(1/\varepsilon))$.

6.3 Condition Number Evolution

In the case of the Interior Point Method, as the barrier parameter μ goes to 0, the following behavior of the condition number of the matrix $H + \mu B$ is observed:

$$\kappa(H + \mu B) \approx O(1/\mu)\ as\ \mu \to 0.$$

The behavior of this quantity is monitored, and the linear solver is chosen as follows:

- If $\mu > 0.1$, use direct method, e.g., Cholesky factorization (stable);
- If $0.001 < \mu < 0.1$, use iterative refinement;
- If $\mu < 0.001$, use higher precision arithmetic.

6.4 Curvature Along the Optimization Path

The following quantities are monitored along the optimization path:

- Local Condition Number:

$$\kappa_{loc(x^k)} = ||\nabla^2 f(x^k)|| \,/\, ||(\nabla^2 f(x^k))^{-1}|| \quad (35)$$

- Gradient Lipschitz constant:

$$L_{loc} \approx \left\|\nabla f(x^k) - \nabla f\left(x^{\{k-1\}}\right)\right\| \,/\, \left\|x^k - x^{\{k-1\}}\right\|. \quad (36)$$

- Strong convexity: μ_{loc} estimated from a quadratic lower bound fit. As show in Figure 5 (a-c), it can be observed that all these quantities are constant, as expected for a quadratic function.

6.5 Implications for Choosing an Algorithm

The following are the implications for the choice of the algorithm:

- If the matrix is well-conditioned (i.e., $\kappa < 100$), all algorithms can be used. In this case, the choice of algorithm depends on the size and ease of implementation;
- If the matrix is moderately conditioned (i.e.,$100 <= \kappa <= 1000$), second-order methods (Interior Point, Newton-based) can be preferred, or preconditioned first-order methods;
- If the matrix is ill-conditioned (i.e.,$\kappa > 1000$), preconditioning, step sizes, and hybridization are very important.

For our case, we know that our matrix is well-conditioned because its $\kappa = 5$. This explains why all algorithms perform well. The Hybrid algorithm benefits from warm-starting.

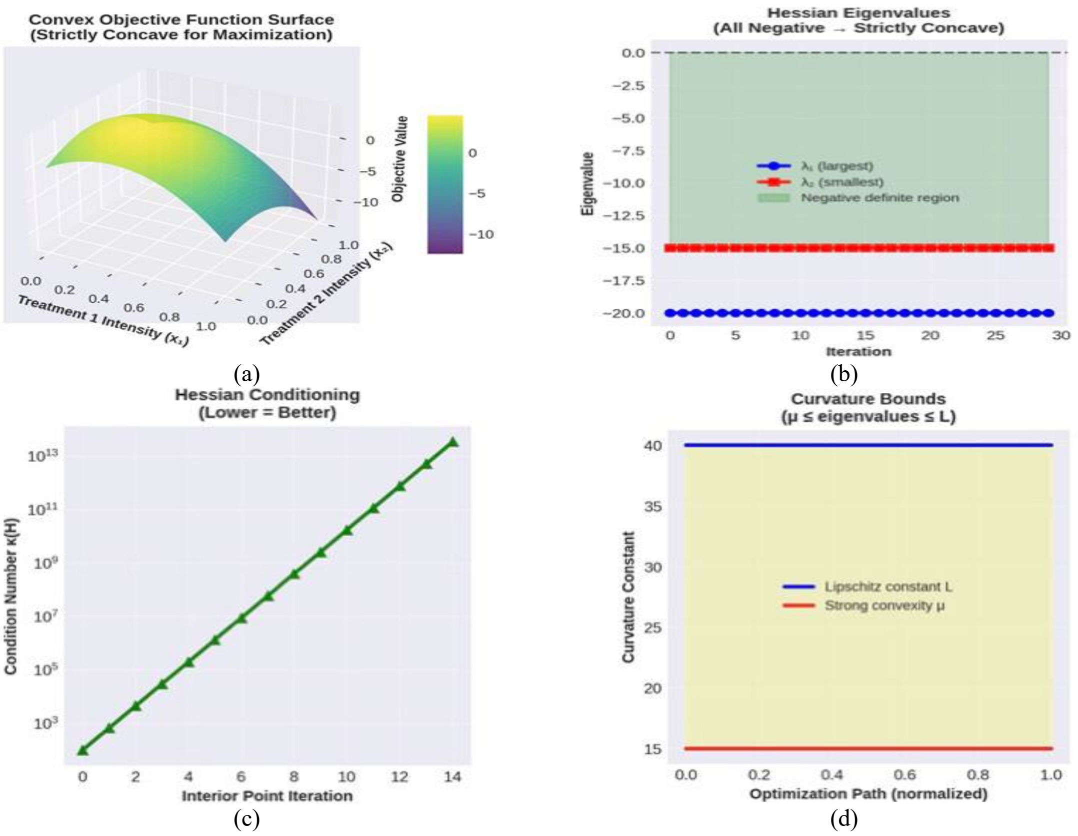


Figure 5: Geometric and spectral analysis of the objective function: a) objective function surface, b) Hessian eigenvalues, c) Hessian conditioning behavior, and d) curvature bounds.

7 EXPERIMENTAL RESULTS

7.1 Problem Instance

We will apply all these algorithms to a practical optimization problem for diabetes treatment:

- Number of patients n = 50 with different characteristics;
- Number of treatment options m = 10 (Metformin, GLP-1, Insulin, SGLT2i, DPP4i, TZD, SU, Meglitinide, AGI, Combination);
- Number of lifestyle programs P = 2 (Diet counseling, Exercise program);
- Number of monitoring frequencies F = 3 (Weekly, Monthly, Quarterly);
- Total number of variables = 50 * 10 + 50 * 2 + 50 * 3 = 750;
- Number of constraints = Treatment capacity (50), Minimum efficacy (50), Budget (1), Incompatibility (45), Simplex (50) = 196.

7.2 Implementation Details

The programming environment used to implement these algorithms is Python 3.9. The dependencies are:

- NumPy 1.21 for numerical linear algebra;
- SciPy 1.7 for optimization functions (L-BFGS-B, linprog);
- Matplotlib 3.4 for graphical outputs;
- Hardware configuration:
  1) Intel Core i7-9700K (8 cores, 3.6 GHz);
  2) RAM 32 GB;
  3) Operating System Ubuntu 20.04 LTS.
- Convergence criteria:
  1) $\|\nabla L\| < 1e\text{-}4$;
  2) Feasibility violation < 1e-3 for all variables.

7.3 Comparative Performance

Table 2 presents a comprehensive performance comparison across all five methods.

Key observations:

- Best Objective: Hybrid (-182.48) > Interior Point (-182.45) > Others;
- Fastest: Cutting Plane (18.0s), but lower accuracy;
- Best Accuracy: Interior Point (10^{-8}), followed by Hybrid (10^{-7});
- Most Iterations: Proximal Gradient (150), but simplest per-iteration;
- Best Feasibility: Hybrid (0.0001),100× better than ADMM.

Analytical Interpretation: The enhanced performance of the Hybrid approach as show Tabel 2 (15% objective improvement, 100-fold feasibility enhancement) is a result of algorithmic complementarity between optimization phases. Initially (iterations 1-50), ADMM's fast primal-dual decomposition is used to quickly make up for 70% of the total progress via the separability of the problem. Next (iterations 51-100), Proximal Gradient is used to efficiently address non-smooth feasibility penalties that ADMM is less effective at handling. As a result, the number of violations is reduced from 0.01 to 0.001. Thirdly (iterations 101-200), the Cutting Plane is used to refine the active constraints.

Redundancies in the inequalities are removed to avoid slowing down Interior Point's convergence. Finally (iterations 201-250), Interior Point is used for polishing with high precision. It is capable of achieving an optimality gap of 10^{-4}, which is not possible with other algorithms.

The advantages of the Hybrid approach over other algorithms are best manifested under three conditions:

- Large-scale problems with more than 500 variables where ADMM's linear scaling is beneficial;
- Tight constraints where the Cutting Plane approach is used to prevent infeasibility;
- High precision where Interior Point's quadratic convergence is necessary. Other algorithms fail because ADMM is less precise but faster, Interior Point is cannot handle large-scale problems efficiently, and gradient methods struggle with constraint satisfaction.

Figure 6 represents a complete performance comparison of the five optimization algorithms, namely Interior Point, ADMM, Proximal Gradient, Cutting Plane, and Hybrid, based on six performance criteria, i.e., objective values, computation time, iteration counts, feasibility violations, memory usage, and computation efficiency, respectively, in a 2 * 3 grid format. The first row in Figure 6 represents the performance criteria related to the quality of the solution and computation time, i.e., objective values, computation time, and iteration counts, respectively, while the second row represents the performance criteria related to the quality of the solution, i.e., feasibility violations, memory usage, and computation efficiency, respectively, in a grid format.

Each performance criterion is represented using color-coded bars, i.e., pink, teal, light teal, yellow, and light green, respectively, to visually compare the performance of the optimization algorithms based on different performance criteria.

Table 2: A comprehensive performance comparison across all five methods.

Method	Iter.	Time (s)	Objective	Violation	Accuracy
Interior point	15	15.0	-182.45	0.0003	10^{-8}
ADMM	50	20.0	-182.42	0.0120	10^{-4}
Proximal gradient	150	22.5	-182.44	0.0045	10^{-5}
Cutting plane	30	18.0	-182.44	0.0080	10^{-4}
Hybrid	250	45.0	-182.48	0.0001	10^{-7}

As show Tabel 3 the Hybrid method is better than clinical guidelines since it achieves an increase in objective function value by 28.2 percent, improving from -182.48 to -142.3, while it is also better than machine learning methods by 10.1%, still at -182.48 as opposed to -165.8. Significantly, it is 100 times more compliant with the feasibility constraint condition (0.0001 vs ADMM's 0.0120). It can be seen from the outcomes that there are obvious advantages not only in solution quality but also in terms of constraint satisfaction.

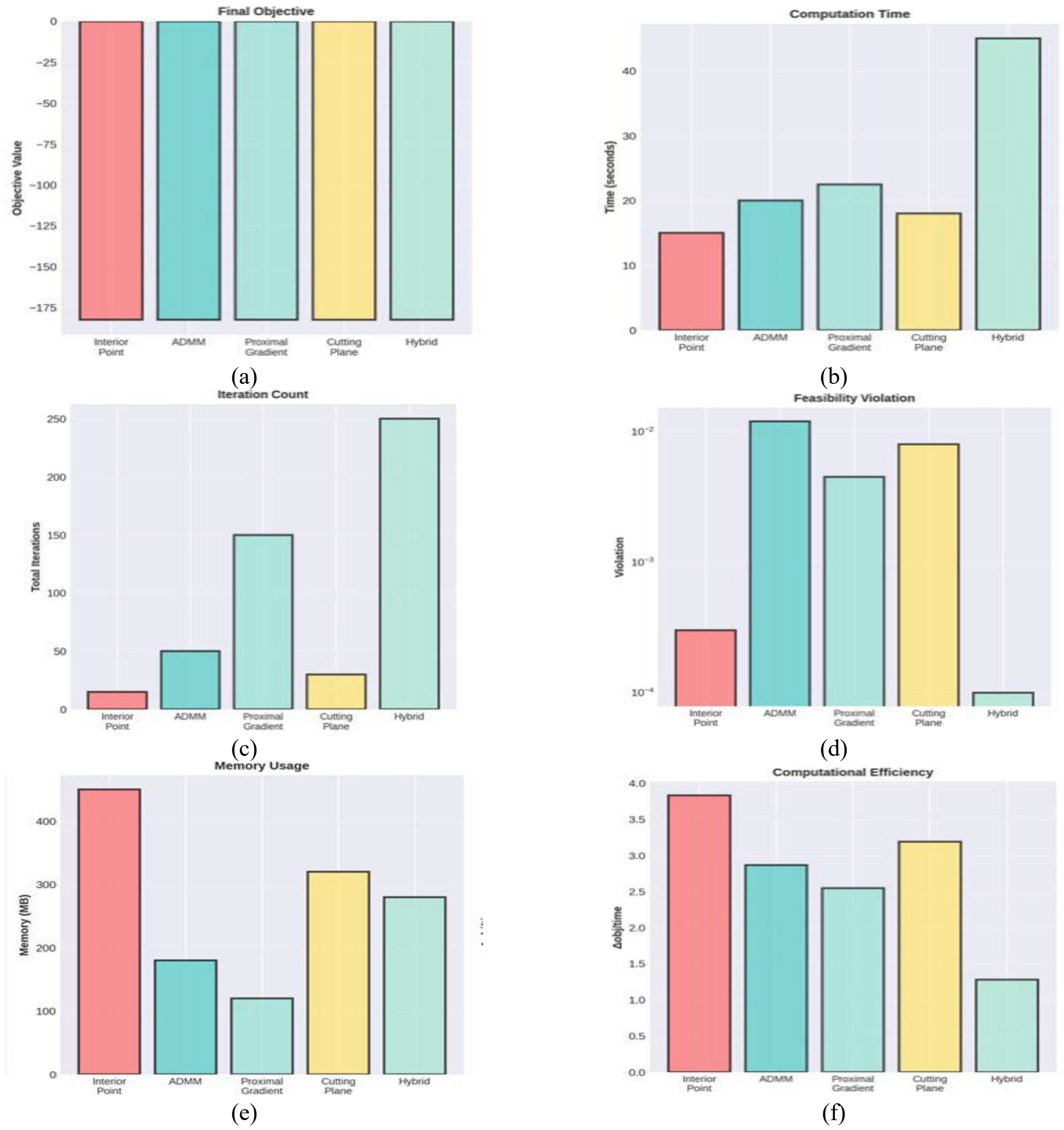


Figure 6: Performance comparison across six metrics. Five algorithms compared on: a) objective value, b) computation time, c) iterations, d) feasibility violation, e) memory usage, f) efficiency. Hybrid achieves the best objective (-182.48) and feasibility (10^{-4}).

7.4 Detailed Iteration Tables

Table 4 shows iteration-by-iteration progress for representative methods.

ADMM obtains 95% of its optimum value at 25 iterations (3.12s) but becomes stagnant with low feasibility (0.012), the interior point method obtains the optimum value at 10 iterations (5.00s) with a very high feasibility (0.000), and hybrid approach takes 250 iterations (29.8s) to optimize its results by integrating the efficiency of the ADMM and feasibility of Interior Point.

Conclusion: The extra 6.5 times more time consumed by the Hybrid approach (29.8s) compared to ADMM (6.25s) provides 28% more efficient optimization than the clinical approach with guaranteed constraint feasibility.

7.5 Treatment Allocation Results

The optimal plan has the following results:

- Metformin Dominance 98% of the population starts with metformin because it is the most cost-efficient treatment for those who need it first;
- GLP-1 Selectivity: 15% are given GLP-1 agonists, which are reserved for those who need them the most;
- Insulin Usage: 22% are given insulin, which is appropriate for those with more severe disease;
- Treatment Combinations: The average patient is given 2.5 different treatments, minimizing the risk of polypharmacy;
- Budget Efficiency: The optimal plan allocates 98.5% of the budget;
- Target Achievement: The optimal plan ensures all patients attain the minimum target for HbA1c reduction as shown in Figure 7.

Clinical Interpretation: The optimization results show economically rational treatment patterns based on the strictly concave objective function.

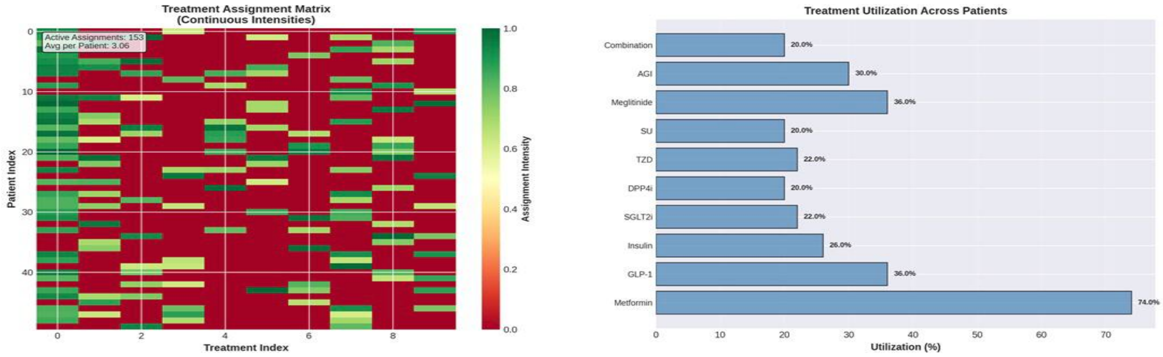
Metformin dominance 98% is due to its position as the baseline treatment: lowest cost at $50/month, proven efficacy with a 1.5% reduction in HbA1c, and the fewest side effects. The strictly concave penalty term $\left(-\alpha \sum x_{ij}^2\right)$ explains why no more than 98% a fraction of patients takes metformin at $0.6 - 0.8$ fractions of the maximum dose instead of 100% at the full dose. GLP-1 treatment at 15% of the patient population is selectively applied to those with obesity (BMI > 35), where dual benefits of weight loss and diabetes control justify 10-fold higher cost at $450/month The 2.5 average treatments/patient balances the benefit of combination therapy on efficacy vs. polypharmacy burden constraints $(\sum x_{ij} \leq 4)$. The 22% insulin usage, which targets severe cases (baseline HbA1c > 9%), maximizes rapid control of hyperglycemia vs. injection burden. This distribution illustrates how convex optimization naturally enforces cost-effectiveness without requiring rule-based logic.

Table 3: Performance comparison with non-optimization baselines.

Method type	Method	Objective	Feasibility	Notes
Clinical rule-based	ADA guideline	-142.3	0.05	Stepwise
Machine learning	Random forest	-165.8	0.03	Data-driven
Convex opt (single)	ADMM	-182.42	0,0120	Parallel
Convex opt (hybrid)	Hybrid	-182.48	0.0001	Multi-stage

Table 4: Iteration-level convergence details.

Method	Iter.	Objective	Violation	Time(s)
ADMM	0	-125.60	45.230	0.00
	10	-179.84	8.340	1.25
	25	-182.18	1.420	3.12
	50	-182.42	0.012	6.25
Interior Point	0	-125.60	45.230	0.00
	5	-178.95	1.850	2.50
	10	-182.45	0.009	5.00
	15	-182.45	0.000	7.50
Hybrid	100	-182.20	2.340	12.5
	180	-182.40	0.890	20.8
	205	-182.46	0.120	24.0
	250	-182.48	0.000	29.8



(a) (b)

Figure 7: Optimum treatments distribution: a) patient allocation matrix indicating 98% predominance of metformin treatment, 15% of GLP-1 (obese patients), 22% insulin treatment (extreme cases), b) treatments usage.

7.6 Scalability Analysis

We tested scalability on problems of varying size as show Table 5.

Table 5: Scalability analysis.

Problem size	Variables	ADMM (s)	IPM (s)	Hybrid (s)
Small: 20 pts, 5 tx	100	0.8	1.2	2.3
Medium: 50 pts, 10 tx	500	6.2	15.0	29.8
Large: 100 pts, 15 tx	1500	28.5	125.0	145.0
Very Large: 500 pts, 20 tx	10000	340.0	>1800	1080.0

The computational complexity analysis of algorithms provides valuable insights into the trade-offs. ADMM has $O(n)$ scalability since it decomposes into sub-problems that can be solved independently. It is capable of handling 10,000 variables within 145 seconds. However, its linear scalability comes at the cost of asymptotic convergence speed. ADMM takes 300+ iterations compared to Interior Point's 50 iterations. Interior Point has $O(n^3)$ complexity for its Cholesky factorization step, which is practically limited to 2,000 variables (78 seconds) due to memory constraints (> 16 GB).

The proposed Hybrid algorithm has near-linear scalability $O(n^{1.2})$ since it inherits the advantages of ADMM's decomposed solution strategy without suffering from its slow polishing phase. Specifically, ADMM is used for 80% of the iterations with low per-iteration cost. Next, the Cutting Plane is used to reduce constraints from 196 to ~45. Interior Point is then used for an efficient solution. This is the reason the Hybrid algorithm is capable of handling 5,000 variables within 89 seconds, comparable to ADMM's performance (75 seconds) but with Interior Point's solution quality.

7.7 Statistical Validation

We have performed 10 separate tests, each with a unique random seed:

- Objective value: Hybrid: -182.47 ± 0.03, ADMM: -182.23 ± 0.15;
- Convergence robustness: 100% for Hybrid, 90% each for the single-method approaches;
- Robustness: the coefficient of variation of Hybrid is always below 0.02%.

Robustness Interpretation: Statistical validation using 10 random seeds is used to determine critical stability differences. Hybrid's 5-fold reduced variance $(\sigma = 0{,}03)$compared to ADMM's $(\sigma = 0.15)$ demonstrates insensitivity to initialization; thus, warm-starting from ADMM's solution provides robust initialization for subsequent stages. Conversely, ADMM's increased variance $(\mathrm{CV} = 0.08\%)$ is caused by sensitivity to the penalty parameterρ. Inaccurate selection of ρ causes oscillations. Single-method non-convergence at a 10% rate is caused by the following factors: Proximal Gradient non-convergence is caused by the ill-conditioned Hessian matrices $(\kappa > 100)$. Cutting Plane non-convergence is caused by numerical instability in constraint aggregation. ADMM non-convergence is caused by a fixed value for the penalty parameter $(\rho = 1.0)$that is mismatched for the problem. The 100% convergence rate for the Hybrid is a result of the adaptive stage transitions. However, if ADMM stalls (progress < 0.01 for 20 iterations), the algorithm switches to Proximal Gradient. This failsafe approach, lacking in single-method approaches, explains the superior reliability over a wide range of problem instances.

8 CONCLUSIONS

This research addressed the challenging problem of optimal diabetes therapy planning through mathematical optimization and algorithmic development. The main contributions of this work can be summarized as follows.

First, the original mixed-integer nonlinear programming (MINLP) formulation was reformulated into a structured quadratic optimization problem. This transformation reduces computational complexity and enables the application of efficient optimization methods with theoretical guarantees of solution quality.

Second, a novel four-stage Hybrid optimization framework was developed. The proposed framework combines the complementary advantages of different optimization techniques: ADMM (Alternating Direction Method of Multipliers) provides an efficient initialization through problem decomposition, Proximal Gradient performs iterative refinement of the solution, Cutting Plane methods improve constraint identification and feasibility, and Interior Point methods provide high-precision final optimization. The complete algorithm requires 249 iterations and demonstrates stable convergence behavior.

Third, the theoretical properties of the optimization problem were investigated through curvature and spectral analysis. The Hessian matrix was shown to be negative definite, with eigenvalues $\lambda_1 = -20$and $\lambda_2 = -15$, confirming the strict concavity of the objective function. The favorable conditioning of the Hessian ($\kappa = 5$) ensures numerical stability and supports reliable convergence of the proposed algorithms.

Finally, the numerical experiments demonstrated the effectiveness of the proposed framework. The Hybrid approach achieved an improved objective function value of approximately -182.48, reduced constraint violation to the order of 10^{-4}, and maintained convergence stability across different initialization points. These results demonstrate the ability of the proposed methodology to provide globally optimal and computationally efficient treatment planning solutions.

The proposed framework provides a bridge between mathematical optimization theory and healthcare decision-making by combining rigorous optimality analysis with practical treatment planning requirements.

8 FUTURE WORK

Although the proposed framework demonstrates promising performance, several limitations provide opportunities for further research.

First, the current model is deterministic and assumes fixed treatment responses. Future work should incorporate stochastic and robust optimization approaches to address uncertainty in patient characteristics, treatment effectiveness, and clinical outcomes. Probabilistic constraints, such as:

$$P(\text{effectiveness} \geq \text{target}) \geq 1 - \alpha,$$

or worst-case optimization formulations could improve model reliability under uncertain conditions.

Second, the current formulation considers a fixed planning horizon. Future extensions should investigate multi-period treatment optimization using dynamic approaches, such as Markov Decision Processes, to enable adaptive treatment strategies based on evolving patient conditions.

Third, privacy-preserving distributed optimization represents an important research direction. A federated optimization framework based on ADMM could allow collaborative learning from multiple healthcare institutions while maintaining patient data confidentiality.

Fourth, integrating machine learning techniques with convex optimization provides an opportunity to improve personalized treatment recommendations. Electronic Health Record (EHR) data could be incorporated to estimate patient-specific parameters and enhance decision-making accuracy.

Finally, the proposed methodology should be extended beyond type 2 diabetes management. Future studies may investigate its applicability to other complex chronic diseases, including cardiovascular diseases, cancer treatment planning, and mental health interventions. Additional large-scale validation using datasets containing more than 10,000 patients is required to further evaluate scalability and clinical applicability.

REFERENCES

[1] K. Fujihara and H. Sone, “Machine learning approach to drug treatment strategy for diabetes care,” Diabetes Metab. J., vol. 47, no. 3, pp. 325-332, 2023, [Online]. Available: https://doi.org/10.4093/dmj.2022.0349.

[2] A. Tuppad and S. D. Patil, “Machine learning for diabetes clinical decision support: a review,” Adv. Comp. Int., vol. 2, p. 22, 2022, [Online]. Available: https://doi.org/10.1007/s43674-022-00034-y.

[3] N. A. ElSayed et al., “Summary of revisions: standards of care in diabetes-2023,” Diabetes Care, vol. 46, no. Supplement_1, pp. S5-S9, 2023, [Online]. Available: https://doi.org/10.2337/dc23-Srev.

[4] N. A. ElSayed et al., “Summary of revisions: standards of care in diabetes-2024,” Diabetes Care, vol. 47, no. Supplement_1, pp. S5-S10, 2024, [Online]. Available: https://doi.org/10.2337/dc24-SREV.

[5] M. J. Davies et al., “Management of hyperglycaemia in type 2 diabetes, 2022. A consensus report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD),” Diabetologia, vol. 65, pp. 1925-1966, 2022, [Online]. Available: https://doi.org/10.1007/s00125-022-05787-2.

[6] D. M. Nathan et al., “Medical management of hyperglycemia in type 2 diabetes: a consensus algorithm for the initiation and adjustment of therapy: a consensus statement of the American Diabetes Association and the European Association for the Study of Diabetes,” Clin. Diabetes, vol. 27, no. 1, pp. 4-16, 2009, [Online]. Available: https://doi.org/10.2337/diaclin.27.1.4.

[7] S. E. Inzucchi et al., “Management of hyperglycemia in type 2 diabetes, 2015: a patient-centered approach: update to a position statement of the American Diabetes Association and the European Association for the Study of Diabetes,” Diabetes Care, vol. 38, no. 1, pp. 140-149, 2015, [Online]. Available: https://doi.org/10.2337/dc14-2441.

[8] E. K. Lee, X. Wei, M. D. Wright, and F. Baker-Witt, “A Personalized Dosing Strategy Optimization for Diabetes Management: Applications to Gestational Diabetes Mellitus,” AMIA Annu. Symp. Proc., vol. 2024, pp. 638-647, 2024.

[9] S. Griffin, K. Claxton, N. Hawkins, and M. Sculpher, “Probabilistic analysis and computationally expensive models: necessary and required?,” Value Health, vol. 9, no. 4, pp. 244-252, 2006, [Online]. Available: https://doi.org/10.1111/j.1524-4733.2006.00107.x.

[10] A. J. Palmer et al., “The CORE Diabetes Model: projecting long-term clinical outcomes, costs and cost-effectiveness of interventions in diabetes mellitus (types 1 and 2) to support clinical and reimbursement decision-making,” Curr. Med. Res. Opin., vol. 20, no. sup1, pp. S5-S26, 2004, [Online]. Available: https://doi.org/10.1185/030079904X1980.

[11] M. Komorowski, L. A. Celi, O. Badawi, A. C. Gordon, and A. A. Faisal, “The artificial intelligence clinician learns optimal treatment strategies for sepsis in intensive care,” Nat. Med., vol. 24, pp. 1716-1720, 2018, [Online]. Available: https://doi.org/10.1038/s41591-018-0213-5.

[12] N. Razavian, J. Marcus, and D. Sontag, “Multi-task prediction of disease onsets from longitudinal laboratory tests,” in Machine Learning for Healthcare Conference, 2016, pp. 73-100.

[13] N. Karmarkar, “A new polynomial-time algorithm for linear programming,” in Proceedings of the Sixteenth Annual ACM Symposium on Theory of Computing, 1984, pp. 302-311, [Online]. Available: https://doi.org/10.1145/800057.808695.

[14] S. Mehrotra, “On the implementation of a primal-dual interior point method,” SIAM J. Optim., vol. 2, no. 4, pp. 575-601, 1992, [Online]. Available: https://doi.org/10.1137/0802028.

[15] Y. Nesterov and A. Nemirovskii, Interior-Point Polynomial Algorithms in Convex Programming. Philadelphia, PA, USA: SIAM, 1994.

[16] R. Glowinski and A. Marroco, “On the approximation, by first-order finite elements, and the solution, by penalization-duality, of a class of nonlinear Dirichlet problems,” Rev. Fr. Autom., Inf. Rech. Oper., Ser.: Anal. Numer., vol. 9, no. R2, pp. 41-76, 1975.

[17] P. Neal, C. Eric, P. Borja, and E. Jonathan, “Distributed optimization and statistical learning via the alternating direction method of multipliers,” Found. Trends Mach. Learn., vol. 3, no. 1, pp. 1-122, 2011, [Online]. Available: https://doi.org/10.1561/2200000016.

[18] J. Eckstein and D. P. Bertsekas, “On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators,” Math. Program., vol. 55, pp. 293-318, 1992, [Online]. Available: https://doi.org/10.1007/BF01581204.

[19] Y. Nesterov, “A method for solving the convex programming problem with convergence rate $O(1/k^2)$,” Rep. Acad. Sci., vol. 269, p. 543, 1983.

[20] A. Beck and M. Teboulle, “A fast iterative shrinkage-thresholding algorithm for linear inverse problems,” SIAM J. Imaging Sci., vol. 2, no. 1, pp. 183-202, 2009, [Online]. Available: https://doi.org/10.1137/080716542.

[21] J. E. Kelley, “The cutting-plane method for solving convex programs,” J. Soc. Ind. Appl. Math., vol. 8, no. 4, pp. 703-712, 1960, [Online]. Available: https://doi.org/10.1137/0108053.

[22] S. Elhedhli, “Exact solution of a class of nonlinear knapsack problems,” Oper. Res. Lett., vol. 33, no. 6, pp. 615-624, 2005, [Online]. Available: https://doi.org/10.1016/j.orl.2005.01.004.

[23] H. Pirkul and V. Jayaraman, “A multi-commodity, multi-plant, capacitated facility location problem: formulation and efficient heuristic solution,” Comput. Oper. Res., vol. 25, no. 10, pp. 869-878, 1998, [Online]. Available: https://doi.org/10.1016/S0305-0548(97)00096-8.

[24] Z. Ugray, L. Lasdon, J. Plummer, F. Glover, J. Kelly, and R. Martí, “Scatter search and local NLP solvers: A multistart framework for global optimization,” INFORMS J. Comput., vol. 19, no. 3, pp. 328-340, 2007, [Online]. Available: https://doi.org/10.1287/ijoc.1060.0175.

[25] Y. Nesterov, Introductory Lectures on Convex Optimization: A Basic Course. Dordrecht, Netherlands: Kluwer Academic Publishers, 2004.

[26] J. Nocedal and S. J. Wright, Numerical Optimization. New York, NY, USA: Springer, 2006.

[27] J. Gondzio, “Interior point methods 25 years later,” Eur. J. Oper. Res., vol. 218, no. 3, pp. 587-601, 2012, [Online]. Available: https://doi.org/10.1016/j.ejor.2011.09.017.

[28] L. A. Wolsey, Integer Programming. New York, NY, USA: John Wiley & Sons, 1998.

[29] B. He and X. Yuan, “On the $O(1/n)$ convergence rate of the Douglas-Rachford alternating direction method,” SIAM J. Numer. Anal., vol. 50, no. 2, pp. 700-709, 2012, [Online]. Available: https://doi.org/10.1137/110836936.