



Technische Universität München

Department of Mathematics



Master's Thesis

Fast Solvers for Batched Constrained Optimization Problems

Anne Christopher

Supervisor: Prof. Felix Brandt

Advisor: Stefan Heidekrüger

Submission Date: 01 August 2020

I hereby declare that this thesis is my own work and that no other sources have been used except those clearly indicated and referenced.

München, 01.08.2020, Anne Christopher

Abstract

An under-explored problem in constrained convex optimization is efficiently solving many convex programs that share the same objective and constraint structure but differ in coefficients in each instance. Such problems appear, for example, in computational game theory and market design. Standard commercial solvers such as Gurobi and CPLEX for convex optimization excel at efficiently computing solutions for a single instance but are not able to leverage synergies to efficiently solve many problems of the same kind. The goal of this thesis is exploring possible solutions to the problem of batched constrained optimization, where individual program instances are usually fairly small but applying standard solvers is prohibitive because of lack of parallelism and unnecessary duplicated instantiation overhead. A major part of this thesis focuses on interior point methods for solving constrained optimization problems. The Primal-Dual Interior-Point Method is exploited in building fast converging algorithms with additional speed-up obtained by using computational power of a GPU.

Zusammenfassung

Bei einer in englischer Sprache verfassten Arbeit muss eine Zusammenfassung in deutscher Sprache vorangestellt werden. Dafür ist hier Platz.

Contents

1	Introduction	1
1.1	Background: Quadratic Programming, Duality Theory and KKT Conditions	1
1.2	Background: LU, LDL^T Factorization	1
1.3	Background: Block Elimination using Schur Complement	1
2	Primal-Dual Interior Point Methods	2
2.1	Newton's Method	2
2.2	Interior Point Methods and PDIPM's	3
2.3	Path-Following Algorithms	4
2.4	Mehrotra's Predictor-Corrector Method	5
2.4.1	Outline of Mehrotra's Predictor Corrector Algorithm	6
2.5	Convergence Analysis of PDIPM's	9
2.6	Complexity Analysis of PDIPM's	9
3	Results	10
4	Conclusion	11
	References	12

Chapter 1

Introduction

- 1.1 Background: Quadratic Programming, Duality Theory and KKT Conditions
- 1.2 Background: LU, LDL^T Factorization
- 1.3 Background: Block Elimination using Schur Complement

Chapter 2

Primal-Dual Interior Point Methods

The primal-dual interior point methods are proven to be a class of algorithms that solves a wide range of optimization problems including linear-programming-problems, convex-quadratic-programming-problems, semi-definite-programming etc.[1]. They fall under the category of interior point methods and employ the Newton's method to solve optimization problems in an efficient manner. The research works that gave rise to the field of *Interior-Point Methods* and successively the *Primal-Dual Methods* started with the publication of Karmarkar's paper [2] which marked an important point in the history of optimization algorithms. This chapter walks through the basic building blocks of the primal-dual interior point methods, its evolution as a path-following algorithm, the Mehrotra's predictor-corrector approach and finally looks at the complexity analysis of these algorithms and compare it with other common optimization algorithms like the simplex method.

2.1 Newton's Method

The Newton's method finds the roots of a function by moving along search directions generated by the the liner approximation of the same function starting from a point in the functions domain. From the Taylor's theorem [3] the linear approximation of a function f which is differentiable at a point x_0 is given by:

$$\bar{f}(x) = f(x_0) + (x - x_0)f'(x_0) \quad (2.1)$$

The Newton Step Δx_n can be obtained by finding the root of this linear approximation of f as $f'(x_n)\Delta x_n = -f(x_n)$. In the matrix system the equivalent representation to obtain the newton step would be $J(x_n)\Delta x_n = -F(x_n)$ where $J(x_n)$ is the Jacobian of $F(x)$ at x_n . These newton steps are the search directions along which the the linear approximation of the function will be iteratively evaluated. The Newton Iterates are hoped to converge to the roots of the function $F(x)$ [3].

The primal-dual interior-point methods, like it's name says, solves the primal and it's corresponding dual problem in each iteration by exploiting the Newton Method. The function which the Newton Method tries to solve is comprised of the Karush-Kuhn-Tucker (KKT) equality conditions (or its modifications). For constrained quadratic problems,

the KKT conditions comprise a set of linear equations which can easily solved by the Newton's method. The underlying idea is that, for any convex optimization problem with differentiable objective and constraint functions, any points that satisfy the KKT conditions are primal and dual optimal and have zero duality gap[4].

2.2 Interior Point Methods and PDIPM's

The interior-point methods get its name from the fact that the iterates generated by this method are points in the interior of the feasible region - i.e, the iterates generated are strictly feasible (or satisfies the inequality constraints strictly). This is in contrast to the Simplex method, which moves along the edges of the feasible region. The primal-dual interior-point methods generates these strictly feasible by solving the equality conditions of the KKT system using Newton Method, restricting the iterates to satisfy the inequality conditions of the KKT system strictly.

Consider a QP(with slack variable s) and its associated KKT conditions for optimality, as mentioned in section 1.1 :

$$\begin{aligned} & \text{minimize} && (1/2)x^T Qx + q^T x \\ & \text{subject to} && Gx + s = h, s \geq 0 \\ & && Ax = b, \end{aligned} \tag{2.2}$$

where $Q \in \mathbf{S}_+^n$, $q \in \mathbf{R}^n$, $G \in \mathbf{R}^{p \times n}$, $h \in \mathbf{R}^p$, $A \in \mathbf{R}^{m \times n}$ and $b \in \mathbf{R}^m$ is the problem data and $x \in \mathbf{R}^n$ and $s \in \mathbf{R}^p$. are the variables. Considering the dual variables as $y \in \mathbf{R}^m$ (associated with the equality constraints) and $z \in \mathbf{R}^p$ (associated with the inequality constraints), the KKT conditions for optimality of this problem are:

- $Qx + q + G^T z + A^T y = 0$ Stationarity
- $z_i s_i = 0, \quad i = 1, \dots, p$ Complementary Slackness
- $Gx + s = h, \quad Ax = b, s \geq 0,$ Primal Feasibility
- $z \geq 0$ Dual Feasibility

The Primal-Dual Interior-Point methods, as mentioned before, find the solution to the primal and dual problem by solving this KKT system using the Newton Method. The function F that the Newton method attempts to solve is comprised of the equality conditions in the KKT system given above.

$$F(x, s, z, y) = \begin{bmatrix} -(A^T y + G^T z + Qx + q) \\ -Sz \\ -(Gx + s - h) \\ -(Ax - b) \end{bmatrix} \tag{2.3}$$

where $S = \text{diag}(s)$.

Now, the update steps can be obtained by the Newton Method as $J(x_n)\Delta x_n = -F(x_n)$, precisely here as:

$$\begin{bmatrix} Q & 0 & G^T & A^T \\ 0 & Z & S & 0 \\ G & I & 0 & 0 \\ A & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta s \\ \Delta z \\ \Delta y \end{bmatrix} = \begin{bmatrix} -(A^T y + G^T z + Qx + q) \\ -Sz \\ -(Gx + s - h) \\ -(Ax - b) \end{bmatrix} \quad (2.4)$$

where $Z = \text{diag}(z)[5]$.

The Jacobian Matrix on the left hand side is obtained by partially differentiating the function $F(x, s, z, y)$ with respect to x, s, z and y respectively.

All primal-dual methods generate iterates (x, s, z, y) that satisfy the KKT inequality conditions strictly, that is, $s > 0$ and $z > 0$. By respecting these conditions, the methods avoid spurious solutions, which are points that satisfy $F(x, s, z, y) = 0$ but not $(s, z) \geq 0$ [1]. Most interior-point methods actually require the iterates to be *strictly feasible*: i.e, each (x, s, z, y) must satisfy the linear equality constraints for the primal and dual problems. We can define \mathcal{F} as the primal-dual *feasible set* and \mathcal{F}^0 as the *strictly feasible set* by:

$$\begin{aligned} \mathcal{F} &= \{(x, s, z, y) | Qx + q + G^T z + A^T y = 0, Gx + s = h, Ax = b, (s, z) \geq 0\} \\ \mathcal{F}^0 &= \{(x, s, z, y) | Qx + q + G^T z + A^T y = 0, Gx + s = h, Ax = b, (s, z) > 0\} \end{aligned}$$

If the current point is strictly feasible, then the terms on the right hand side of equation 2.4 vanishes to zero, except the term $-Sz$. However, since it is difficult to find a strictly feasible starting point, the *infeasible interior-point* methods does not require the starting points to belong to the set \mathcal{F}^0 , rather it only requires the positivity of s and z , i.e., $(s, z) > 0$ and solves the system of equations in equation 2.4.

Once the Newton's step is obtained, instead of taking a full step along this newton direction, the best possible *step-size* that can be made along this step direction without violating the inequality condition $(s, z > 0)$ is found using a line search. The updates are then performed as:

$$(x^{k+1}, s^{k+1}, z^{k+1}, y^{k+1}) = (x^k, s^k, z^k, y^k) + \alpha * (\Delta x^k, \Delta s^k, \Delta z^k, \Delta y^k) \quad (2.5)$$

for some line search parameter $\alpha \in (0, 1]$.

Unfortunately, the steps that can be taken along this pure Newton-Directions are very small before violating the inequality condition, and hence makes very little progress towards the solution [1]. This is taken care of by biasing the the search direction - which is discussed in the following section.

2.3 Path-Following Algorithms

Primal-Dual Interior-Point Methods modify the basic Newton's method by biasing the search direction towards the interior of the search space (the non-negative orthant $(s, z) \geq 0$). This is obtained by slightly modifying the KKT conditions by introducing a term τ

in the *complementary slackness* condition.

$$x_i s_i = \tau \quad (2.6)$$

The points in $\mathcal{C} = (x_\tau, s_\tau, z_\tau, y_\tau) | \tau > 0$ is called the **Central Path**. This modified KKT system would approximate the original KKT system as τ goes to zero. Also as $\tau \rightarrow 0$, if \mathcal{C} converges to anything, then this is the optimal - primal-dual solution. The central path drives us to the solution at a much faster rate than the basic Newton Method, since the steps taken are longer in the former compared to the later. The biased search direction τ is defined by using a *centering parameter* $\sigma \in [0, 1]$ and a *duality measure* μ defined as given below: [1].

$$\mu = s^T z / p \quad (2.7)$$

Therefore, the modified KKT system is given by:

$$\begin{bmatrix} Q & 0 & G^T & A^T \\ 0 & Z & S & 0 \\ G & I & 0 & 0 \\ A & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta s \\ \Delta z \\ \Delta y \end{bmatrix} = \begin{bmatrix} -(A^T y + G^T z + Qx + q) \\ -Sz + \sigma \mu e \\ -(Gx + s - h) \\ -(Ax - b) \end{bmatrix} \quad (2.8)$$

When $\sigma = 0$ the above system of equations resembles equation 2.4. The steps then generated are the Standard Newton step often called the *affine-scaling direction*. When $\sigma = 1$ the equation 2.8 defines a *centering direction*. This gives points on the central path, but they help very little in reducing μ significantly. However they set the scene for major progress in the next iteration. To achieve the twin goals of reducing μ and improving centrality, many algorithms use values of σ in the interval $(0,1)$ [1].

These algorithms that restrict the iterates to a neighborhood of the central path \mathcal{C} and follows the central path to reach the solution are called **Path-Following Algorithms**. Some of the common path following algorithms include *short-step* path following methods, *Long-step* path following methods, *predictor-corrector* method etc. The Mehrotra's Predictor Corrector algorithm [6] is one such class of algorithms which has shown consistent performance in solving a wide range of problems. This method is described in the following section.

2.4 Mehrotra's Predictor-Corrector Method

The Mehrotra's predictor-corrector method[6] generates iterates which satisfy the positivity conditions, but might not belong to the strictly feasible set \mathcal{F}^0 mentioned before. This method is characterized by the following three components[1] :

- a pure newton step or an affine-scaling "predictor" direction
- a centering parameter σ which is adaptively chosen based on the affine-newton step
- a "corrector" direction that compensate for some of the non-linearity in the affine-scaling direction.

Unlike the standard path-following primal-dual methods 2.8, Mehrotra's Predictor Corrector Method computes the affine-scaling direction 2.4 and the centering term $\sigma\mu e$ separately. This is done so, because the centering parameter σ is chosen adaptively based on the improvement the affine-scaling direction (generated by the pure newton method) offers. If the affine-scaling direction makes a good progress in reducing the duality measure μ without violating the inequality bounds $(s, z) > 0$, then there is no much centering required, and so σ can be set close to zero for this iteration. On the other hand, if very little progress can be made along the affine-scaling direction, the centering parameter σ is set close to 1 so as to drive the iterates closer to the central path so that they can make significant progress through the next iterations. One disadvantage of splitting the system into two systems is that, two linear systems are to be solved in each iteration. But however, this can be done efficiently, since the coefficient matrix is the same in both the cases and needs to be factorized only once.

The affine-scaling directions are obtained from a linear approximation to the equality conditions in the KKT system as described by equation 2.1. The affine-scaling direction can be used to assess the error in the linear approximation. This error is used to calculate a *corrector* component - this improves the linear, first-order model of F to a quadratic, second-order model.

2.4.1 Outline of Mehrotra's Predictor Corrector Algorithm

Solving the KKT system

It is important to note that the coefficient matrix in the KKT system has a special structure. If we have a closer look at the coefficient matrix of the KKT system 2.4, we realize that, all the matrix entries except Z and S (in the second row) remain unchanged across iterations. This can be utilized to improve the efficiency of the solve steps of the KKT system. This matrix can be symmetrized by dividing the whole second row to obtain the following:

$$\left[\begin{array}{cc|cc} Q & 0 & G^T & A^T \\ 0 & S^{-1}Z & I & 0 \\ \hline G & I & 0 & 0 \\ A & 0 & 0 & 0 \end{array} \right] \begin{bmatrix} \Delta x \\ \Delta s \\ \Delta z \\ \Delta y \end{bmatrix} = \begin{bmatrix} -(A^T y + G^T z + Qx + q) \\ -z \\ -(Gx + s - h) \\ -(Ax - b) \end{bmatrix} \quad (2.9)$$

where $S^{-1}Z$ is a diagonal matrix, represented as $D = \text{diag}(d)$ where $d = z/s$.

In each iteration of the Mehrotra's Predictor-Corrector algorithm, systems similar to the above with same coefficient matrix on the left hand side but different right hand side would be solved. A solver method that efficiently does this solving procedure by a suitable factorization of the coefficient matrix (once per iteration) would help improve the performance of the algorithm.

Initialization

Mehrotra's predictor corrector algorithm works with infeasible iterates and only requires the positivity condition $(s, z) > 0$. Hence an initial point can be generated in many ways, one among which is given below as described in [7]. This method tries to find the

(analytic) solution of the primal and dual problems :

$$\begin{aligned} & \text{minimize} && (1/2)x^T Qx + q^T x + (1/2)\|s\|_2^2 \\ & \text{subject to} && Gx + s = h, \quad Ax = b \end{aligned}$$

with variables x and s , and

$$\begin{aligned} & \text{maximize} && -(1/2)w^T Qw - h^T z - b^T y - (1/2)\|z\|_2^2 \\ & \text{subject to} && Qw + q + G^T z + A^T y = 0 \end{aligned}$$

with variables w, y and z . The solution to the KKT system of these problems would reduce to:

$$\begin{bmatrix} Q & G^T & A^T \\ G & -I & 0 \\ A & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ z \\ y \end{bmatrix} = \begin{bmatrix} -q \\ h \\ b \end{bmatrix}$$

If we have a solver to solve the general KKT system across iterations, we can use that solver to find solutions to the problems above by passing in the matrix system given below which resembles the structure of 2.10.

$$\begin{bmatrix} Q & 0 & G^T & A^T \\ 0 & I & I & 0 \\ G & I & 0 & 0 \\ A & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta s \\ \Delta z \\ \Delta y \end{bmatrix} = \begin{bmatrix} -q \\ 0 \\ h \\ b \end{bmatrix} \quad (2.10)$$

The initial primal and dual variables x and y are set to $x^{(0)} = x$ and $y^{(0)} = y$. However, we require the positivity of the variable s and z . For achieving this, we set $z = Gx - h$ and $\alpha_p = \inf\{\alpha \mid -z + \alpha \mathbf{1} \geq 0\}$, and use

$$s^{(0)} = \begin{cases} -z & \alpha_p < 0 \\ -z + (1 + \alpha_p) \mathbf{1} & \text{otherwise} \end{cases}$$

as the initial value of s . Finally, we set $\alpha_d = \inf\{\alpha \mid z + \alpha \mathbf{1} \geq 0\}$, and use

$$z^{(0)} = \begin{cases} z & \alpha_d < 0 \\ z + (1 + \alpha_d) \mathbf{1} & \text{otherwise} \end{cases}$$

as the initial value of z . This gives the starting point $(x^{(0)}, s^{(0)}, z^{(0)}, y^{(0)})$.

Main Iterations

1. The optimum can be said to be achieved once the residuals and the duality gap approaches zero. This can be used as a early stopping criterion, for avoiding unnecessary computations until max_iteration is reached. (The residuals are the difference of the KKT equality conditions from zero. Precisely $-(A^T y + G^T z + Qx + q)$ is the dual residual and $-(Gx + s - h) - (Ax - b)$ form the primal residual.)

2. The affine-scaling directions are found using pure-newton step as:

$$\begin{bmatrix} Q & 0 & G^T & A^T \\ 0 & Z & S & 0 \\ G & I & 0 & 0 \\ A & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta x^{aff} \\ \Delta s^{aff} \\ \Delta z^{aff} \\ \Delta y^{aff} \end{bmatrix} = \begin{bmatrix} -(A^T y + G^T z + Qx + q) \\ -Sz \\ -(Gx + s - h) \\ -(Ax - b) \end{bmatrix} \quad (2.11)$$

3. Using these affine-scaling directions obtained above, the centering parameter σ is calculated as:

$$\sigma = \left(\frac{(s + \alpha \Delta s^{\text{aff}})^T (z + \alpha \Delta z^{\text{aff}})}{s^T z} \right)^3 = \left(\frac{\mu^{\text{aff}}}{\mu} \right)^3 \quad (2.12)$$

where,

$$\alpha = \sup \{ \alpha \in [0, 1] | s + \alpha \Delta s^{\text{aff}} \geq 0, z + \alpha \Delta z^{\text{aff}} \geq 0 \} \quad (2.13)$$

i.e, the centering parameter is adaptively chosen based on the improvement offered by the affine-scaling direction alone. μ^{aff} refers to the new duality measure if only the affine-scaling direction was used for update. Also the duality measure μ is calculated as $\mu = s^T z / p$.

4. The centering parameter and the duality measure can be used to determine the centering term that biases the search direction. We can combine this centering step with the corrector step that accounts for the non-linearity in the affine-step. The combined centering-corrector steps are obtained by solving the following system:

$$\begin{bmatrix} Q & 0 & G^T & A^T \\ 0 & Z & S & 0 \\ G & I & 0 & 0 \\ A & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta x^{ce} \\ \Delta s^{cc} \\ \Delta z^{cc} \\ \Delta y^{cc} \end{bmatrix} = \begin{bmatrix} 0 \\ \sigma \mu 1 - \text{diag}(\Delta s^{\text{aff}}) \Delta z^{\text{aff}} \\ 0 \\ 0 \end{bmatrix} \quad (2.14)$$

5. Once the affine-steps and the centering-corrector-steps are obtained, the update-steps can be calculated as:

$$\begin{aligned} \Delta x &= \Delta x^{2/f} + \Delta x^{cc} \\ \Delta s &= \Delta s^{\text{aff}} + \Delta s^{cc} \\ \Delta y &= \Delta y^{\text{aff}} + \Delta y^{cc} \\ \Delta z &= \Delta z^{\text{aff}} + \Delta z^{\infty} \end{aligned} \quad (2.15)$$

6. Finally we perform the updates of the variables as:

$$\begin{aligned} x &:= x + \alpha \Delta x \\ s &:= s + \alpha \Delta s \\ y &:= y + \alpha \Delta y \\ z &:= z + \alpha \Delta z \end{aligned} \quad (2.16)$$

where

$$\alpha = \min\{1, 0.99 \sup\{\alpha \geq 0 | s + \alpha \Delta s \geq 0, z + \alpha \Delta z \geq 0\}\} \quad (2.17)$$

6. Repeat this steps from 1. until max_iteration is reached

2.5 Convergence Analysis of PDIPM's

2.6 Complexity Analysis of PDIPM's

Chapter 3

Results

Chapter 4

Conclusion

References

- [1] Stephen J Wright. *Primal-dual interior-point methods*, volume 54. Siam, 1997.
- [2] Narendra Karmarkar. A new polynomial-time algorithm for linear programming. In *Proceedings of the sixteenth annual ACM symposium on Theory of computing*, pages 302–311, 1984.
- [3] Tom M Apostol. *Mathematical analysis*. 1964.
- [4] Stephen Boyd, Stephen P Boyd, and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.
- [5] Jacob Mattingley and Stephen Boyd. Cvxgen: A code generator for embedded convex optimization. *Optimization and Engineering*, 13(1):1–27, 2012.
- [6] Sanjay Mehrotra. On the implementation of a primal-dual interior point method. *SIAM Journal on optimization*, 2(4):575–601, 1992.
- [7] Lieven Vandenberghe. The cvxopt linear and quadratic cone program solvers. *Online: <http://cvxopt.org/documentation/coneprog.pdf>*, 2010.