A Tutorial on Convex Optimization II: Duality and Interior Point Methods

Haitham Hindi

Palo Alto Research Center (PARC), Palo Alto, California 94304
email: hhindi@parc.com

Abstract— In recent years, convex optimization has become a computational tool of central importance in engineering, thanks to its ability to solve very large, practical engineering problems reliably and efficiently. The goal of this tutorial is to continue the overview of modern convex optimization from where our ACC2004 Tutorial on Convex Optimization left off, to cover important topics that were omitted there due to lack of space and time, and highlight the intimate connections between them. The topics of duality and interior point algorithms will be our focus, along with simple examples. The material in this tutorial is excerpted from the recent book on convex optimization, by Boyd and Vandenberghe, who have made available a large amount of free course material and freely available software. These can be downloaded and used immediately by the reader both for self-study and to solve real problems.

I. INTRODUCTION

The objectives are to continue the overview of modern convex optimization from where our ACC2004 Tutorial on Convex Optimization [18] left off. Specifically, we review the role of duality and demonstrate both its practical and theoretical value in convex optimization; describe interior point methods for solving convex optimization problems; and highlight the intimate connections between duality and the solution methods.

We aim to give an overview of the essential ideas mainly defining concepts and listing properties without proof. With the exception of minimal narrative comments by the present author, all of the material in this tutorial is excerpted from chapters 5, 9, 10 and 11 of the book Convex Optimization by Boyd and Vandenberghe [8], where complete details with lucid explanations can be found. This will be our main reference in this tutorial. I am deeply indebted to the authors, for generously allowing me to use their material in preparing this tutorial. The authors have also made available on the internet a large amount of free course material and software [14], [22].

The reader is assumed to have a working knowledge of linear algebra and basic vector calculus, and some (minimal) exposure to optimization. However, due to its different focus, this paper can be read quite independently of our Part I paper [18]. The following references also cover the topics of optimization [26], [24], [1], [3], [4], convex analysis [23], [28], [30], [19], [15], and numerical computation [29], [13], [11], [20].

Also, in order keep the paper quite general, we have tried to not to bias our presentation toward any particular audience. Hence, the examples used in the paper are very simple and intended merely to clarify the optimization ideas and concepts. For detailed examples and applications, the reader is refered to [8], [2], [6], [5], [7], [10], [12], [17], [9], [25], [16], [31], and the references therein.

We now briefly outline the paper. There are two main sections after this one. Section II is on duality, where we summarize the key ideas the general theory, illustrating the four main practical applications of duality with simple examples. Section III is on interior point algorithms, where the focus is on barrier methods, which can be implemented easily using only a few key technical components, and yet are highly effective both in theory and in practice. All of the theory we cover can be readily extended to general conic programs, such as second order cone programs (SOCP) and semidefinite programs (SDP); see [8] for details.

Notation Our notation is standard [8]. For example, we will use **dom** to denote the domain of a function, **int** denotes the interior of a set, **relint** denotes the interior of a set relative to the smallest affine set containing it (its affine hull), and $\succeq$ ($\preceq$) to denote componentwise inequality when applied to vectors, and semidefinite ordering, when applied to matrices. $\mathbf{S}^n_+$ denotes the set of symmetric $n \times n$ positive semidefinite matrices.

II. DUALITY

In this section, we present the basic ideas of duality theory, illustrating along the way its four main practical uses: bounding nonconvex problems; stopping criteria in algorithms; decomposition of large problems; and sensitivity analysis.

A. The Lagrange dual function

1) The Lagrangian: We consider an optimization problem in the standard form:

$$
\begin{aligned}
\text{minimize} \quad & f_0(x) \\
\text{subject to} \quad & f_i(x) \leq 0, \quad i = 1, \ldots, m \\
& h_i(x) = 0, \quad i = 1, \ldots, p,
\end{aligned} \tag{1}
$$

with variable $x \in \mathbf{R}^n$, and where the functions from $\mathbf{R}^n \to \mathbf{R}$ $f_0, f_1, \ldots, f_m$, and $h_1, \ldots, h_p$ are the objective, inequality constraints and equality constraints, respectively. We assume its domain $\mathcal{D} = \bigcap_{i=0}^m \mathbf{dom}\, f_i \cap \bigcap_{i=1}^p \mathbf{dom}\, h_i$ is nonempty, and denote the optimal value of (1) by $p^\star$. For now, we do not assume the problem (1) is convex.

The basic idea in Lagrangian duality is to take the constraints in (1) into account by augmenting the objective function with a weighted sum of the constraint functions. We

define the *Lagrangian* $L : \mathbf{R}^n \times \mathbf{R}^m \times \mathbf{R}^p \to \mathbf{R}$ associated with the problem (1) as

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x),$$

with $\mathbf{dom}\, L = \mathcal{D} \times \mathbf{R}^m \times \mathbf{R}^p$. We refer to λ_i as the *Lagrange multiplier* associated with the ith inequality constraint $f_i(x) \leq 0$; similarly we refer to ν_i as the Lagrange multiplier associated with the ith equality constraint $h_i(x) = 0$. The vectors λ and ν are called the *dual variables* or *Lagrange multiplier vectors* associated with the problem (1).

2) The Lagrange dual function: We define the *Lagrange dual function* (or just *dual function*) $g : \mathbf{R}^m \times \mathbf{R}^p \to \mathbf{R}$ as the minimum value of the Lagrangian over x: for $\lambda \in \mathbf{R}^m$, $\nu \in \mathbf{R}^p$,

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x) \right).$$

When the Lagrangian is unbounded below in x, the dual function takes on the value $-\infty$. Since the dual function is the pointwise infimum of a family of affine functions of (λ, ν), it is concave, even when the problem (1) is not convex.

3) Lower bounds on optimal value: It is easy to show [8] that the dual function yields lower bounds on the optimal value $p^\star$ of the problem (1): For any $\lambda \succeq 0$ and any ν we have

$$g(\lambda, \nu) \leq p^\star. \tag{2}$$

This important property is easily verified. Suppose $\tilde{x}$ is a feasible point for the problem (1), *i.e.*, $f_i(\tilde{x}) \leq 0$ and $h_i(\tilde{x}) = 0$, and $\lambda \succeq 0$. Then we have

$$\sum_{i=1}^{m} \lambda_i f_i(\tilde{x}) + \sum_{i=1}^{p} \nu_i h_i(\tilde{x}) \leq 0,$$

since each term in the first sum is nonpositive, and each term in the second sum is zero, and therefore

$$L(\tilde{x}, \lambda, \nu) = f_0(\tilde{x}) + \sum_{i=1}^{m} \lambda_i f_i(\tilde{x}) + \sum_{i=1}^{p} \nu_i h_i(\tilde{x}) \leq f_0(\tilde{x}).$$

Hence

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) \leq L(\tilde{x}, \lambda, \nu) \leq f_0(\tilde{x}).$$

Since $g(\lambda, \nu) \leq f_0(\tilde{x})$ holds for every feasible point $\tilde{x}$, the inequality (2) follows.

The inequality (2) holds, but is vacuous, when $g(\lambda, \nu) = -\infty$. The dual function gives a nontrivial lower bound on $p^\star$ only when $\lambda \succeq 0$ and $(\lambda, \nu) \in \mathbf{dom}\, g$, *i.e.*, $g(\lambda, \nu) > -\infty$. We refer to a pair (λ, ν) with $\lambda \succeq 0$ and $(\lambda, \nu) \in \mathbf{dom}\, g$ as *dual feasible*, for reasons that will become clear later.

Example *Standard form LP* Consider an LP in standard form,

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & x \succeq 0, \end{array} \tag{3}$$

which has inequality constraint functions $f_i(x) = -x_i$, $i = 1, \ldots, n$. To form the Lagrangian we introduce multipliers λ_i for the n inequality constraints and multipliers ν_i for the equality constraints, and obtain

$$\begin{aligned} L(x, \lambda, \nu) &= c^T x - \sum_{i=1}^{n} \lambda_i x_i + \nu^T (Ax - b) \\ &= -b^T \nu + (c + A^T \nu - \lambda)^T x. \end{aligned}$$

The dual function is

$$g(\lambda, \nu) = \inf_{x} L(x, \lambda, \nu) = -b^T \nu + \inf_{x} (c + A^T \nu - \lambda)^T x,$$

which is easily determined analytically, since a linear function is bounded below only when it is identically zero. Thus, $g(\lambda, \nu) = -\infty$ except when $c + A^T \nu - \lambda = 0$, in which case it is $-b^T \nu$:

$$g(\lambda, \nu) = \begin{cases} -b^T \nu & A^T \nu - \lambda + c = 0 \\ -\infty & \text{otherwise}. \end{cases}$$

Note that the dual function g is finite only on a proper affine subset of $\mathbf{R}^m \times \mathbf{R}^p$. We will see that this is a common occurrence. The lower bound property (2) is nontrivial only when λ and ν satisfy $\lambda \succeq 0$ and $A^T \nu - \lambda + c = 0$. When this occurs, $-b^T \nu$ is a lower bound on the optimal value of the LP (3).

B. The Lagrange dual problem

For each pair (λ, ν) with $\lambda \succeq 0$, the Lagrange dual function gives us a lower bound on the optimal value $p^\star$ of the optimization problem (1). Thus we have a lower bound that depends on some parameters λ, ν. A natural question is: What is the *best* lower bound that can be obtained from the Lagrange dual function?

This leads to the optimization problem

$$\begin{array}{ll} \text{maximize} & g(\lambda, \nu) \\ \text{subject to} & \lambda \succeq 0. \end{array} \tag{4}$$

This problem is called the *Lagrange dual problem* associated with the problem (1). In this context the original problem (1) is sometimes called the *primal problem*. The term *dual feasible*, to describe a pair (λ, ν) with $\lambda \succeq 0$ and $g(\lambda, \nu) > -\infty$, now makes sense. It means, as the name implies, that (λ, ν) is feasible for the dual problem (4). We refer to $(\lambda^\star, \nu^\star)$ as *dual optimal* or *optimal Lagrange multipliers* if they are optimal for the problem (4).

The Lagrange dual problem (4) is a convex optimization problem, since the objective to be maximized is concave and the constraint is convex. This is the case whether or not the primal problem (1) is convex.

1) Making dual constraints explicit: The example above shows that it can happen (and often does) that the domain of the dual function,

$$\mathbf{dom}\, g = \{(\lambda, \nu) \mid g(\lambda, \nu) > -\infty\},$$

to have dimension smaller than $m + p$. In many cases we can identify the affine hull of $\mathbf{dom}\, g$, and describe it as a set of linear equality constraints. Roughly speaking, this means we can identify the equality constraints that are 'hidden' or 'implicit' in the objective g of the dual problem (4). In this case we can form an equivalent problem, in which these equality constraints are given explicitly as constraints. The following examples demonstrate this idea.

Example *Lagrange dual of standard form LP* Continuing the LP example above, the Lagrange dual problem of the standard form LP is to maximize this dual function g subject to $\lambda \succeq 0$, *i.e.*,

$$\text{maximize} \quad g(\lambda, \nu) = \begin{cases} -b^T\nu & A^T\nu - \lambda + c = 0 \\ -\infty & \text{otherwise} \end{cases} \tag{5}$$
$$\text{subject to} \quad \lambda \succeq 0.$$

Here g is finite only when $A^T\nu - \lambda + c = 0$. We can form an equivalent problem by making these equality constraints explicit:

$$\begin{aligned} \text{maximize} \quad & -b^T\nu \\ \text{subject to} \quad & A^T\nu - \lambda + c = 0 \\ & \lambda \succeq 0. \end{aligned} \tag{6}$$

This problem, in turn, can be expressed as

$$\begin{aligned} \text{maximize} \quad & -b^T\nu \\ \text{subject to} \quad & A^T\nu + c \succeq 0, \end{aligned} \tag{7}$$

which is an LP in inequality form, since λ can be viewed as a slack variable.

Note the subtle distinctions in the LP examples so far. The Lagrange dual of the standard form LP (3) is the problem (5), which is equivalent to (but not the same as) the problems (6) and (7). With some abuse of terminology, we refer to the problem (6) or the problem (7) as the Lagrange dual of the standard form LP (3).

2) Weak duality: The optimal value of the Lagrange dual problem, which we denote $d^\star$, is, by definition, the best lower bound on $p^\star$ that can be obtained from the Lagrange dual function. In particular, we have the simple but important inequality

$$d^\star \leq p^\star, \tag{8}$$

which holds even if the original problem is not convex. This property is called *weak duality*.

The weak duality inequality (8) holds when $d^\star$ and $p^\star$ are infinite. For example, if the primal problem is unbounded below, so that $p^\star = -\infty$, we must have $d^\star = -\infty$, *i.e.*, the Lagrange dual problem is infeasible. Conversely, if the dual problem is unbounded above, so that $d^\star = \infty$, we must have $p^\star = \infty$, *i.e.*, the primal problem is infeasible.

We refer to the difference $p^\star - d^\star$ as the *optimal duality gap* of the original problem, since it gives the gap between the optimal value of the primal problem and the best (*i.e.*, greatest) lower bound on it that can be obtained from the Lagrange dual function. The optimal duality gap is always nonnegative.

The bound (8) can sometimes be used to find a lower bound on the optimal value of a problem that is difficult to solve, since the dual problem is always convex, and in many cases can be solved efficiently, to find $d^\star$.

3) Strong duality and Slater's constraint qualification: If the equality

$$d^\star = p^\star \tag{9}$$

holds, *i.e.*, the optimal duality gap is zero, then we say that *strong duality* holds. This means that the best bound that can be obtained from the Lagrange dual function is tight.

Strong duality does not, in general, hold. But if the primal problem (1) is convex, *i.e.*, of the form

$$\begin{aligned} \text{minimize} \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) \leq 0, \quad i = 1, \dots, m, \\ & Ax = b, \end{aligned} \tag{10}$$

with $f_0, \dots, f_m$ convex, we usually (but not always) have strong duality. There are many results that establish conditions on the problem, beyond convexity, under which strong duality holds. These conditions are called *constraint qualifications*.

One simple constraint qualification is *Slater's condition*: There exists an $x \in \mathbf{relint}\,\mathcal{D}$ such that

$$f_i(x) < 0, \quad i = 1, \dots, m, \qquad Ax = b. \tag{11}$$

Such a point is sometimes called *strictly feasible*, since the inequality constraints hold with strict inequality. Slater's theorem states that strong duality holds, if Slater's condition holds (and the problem is convex).

Slater's condition can be refined when some of the inequality constraint functions f_i are affine [8]. Slater's condition (and its refinement) not only implies strong duality for convex problems. It also implies that the dual optimal value is attained when $d^\star > -\infty$, *i.e.*, there exists a dual feasible $(\lambda^\star, \nu^\star)$ with $g(\lambda^\star, \nu^\star) = d^\star = p^\star$. Slater's condition is proved using the separating hyperplane theorem for convex sets, see [8].

C. Optimality conditions

We remind the reader that we do not assume the problem (1) is convex, unless explicitly stated.

1) Certificate of suboptimality and stopping criteria: If we can find a dual feasible (λ, ν), we establish a lower bound on the optimal value of the primal problem: $p^\star \geq g(\lambda, \nu)$. Thus a dual feasible point (λ, ν) provides a *proof* or *certificate* that $p^\star \geq g(\lambda, \nu)$. Strong duality means there exist arbitrarily good certificates.

Dual feasible points allow us to bound how suboptimal a given feasible point is, without knowing the exact value of $p^\star$. Indeed, if x is primal feasible and (λ, ν) is dual feasible, then

$$f_0(x) - p^\star \leq f_0(x) - g(\lambda, \nu).$$

In particular, this establishes that x is ϵ-suboptimal, with $\epsilon = f_0(x) - g(\lambda, \nu)$. (It also establishes that (λ, ν) is ϵ-suboptimal for the dual problem.) We refer to the gap between primal and dual objectives, $f_0(x) - g(\lambda, \nu)$ as the *duality gap* associated with the primal feasible point x and dual feasible point (λ, ν). If the duality gap of the primal dual feasible pair x, (λ, ν) is zero, *i.e.*, $f_0(x) = g(\lambda, \nu)$, then x is primal optimal and (λ, ν) is dual optimal.

These observations can be used in optimization algorithms to provide nonheuristic stopping criteria, as we will see in section III. Suppose an algorithm produces a sequence of primal feasible $x^{(k)}$ and dual feasible $(\lambda^{(k)}, \nu^{(k)})$, for $k = 1, 2, \dots$, and $\epsilon_{\text{abs}} > 0$ is a given required absolute accuracy.

Then the stopping criterion (*i.e.*, the condition for terminating the algorithm)

$$f_0(x^{(k)}) - g(\lambda^{(k)}, \nu^{(k)}) \le \epsilon_{\text{abs}}$$

guarantees that when the algorithm terminates, $x^{(k)}$ is ϵ_{abs}-suboptimal. (Of course strong duality must hold if this method is to work for arbitrarily small tolerances ϵ_{abs}.)

2) Complementary slackness: Suppose that the primal and dual optimal values are attained and equal (so, in particular, strong duality holds). Let $x^\star$ be a primal optimal and $(\lambda^\star, \nu^\star)$ be a dual optimal point. Then it can be shown that the following condition must hold [8]:

$$\lambda_i^\star f_i(x^\star) = 0, \quad i = 1, \dots, m. \tag{12}$$

This condition is known as *complementary slackness*; it holds for any primal optimal $x^\star$ and any dual optimal $(\lambda^\star, \nu^\star)$ (when strong duality holds). We can express the complementary slackness condition as

$$\lambda_i^\star > 0 \implies f_i(x^\star) = 0,$$

or, equivalently,

$$f_i(x^\star) < 0 \implies \lambda_i^\star = 0.$$

Roughly speaking, this means the ith optimal Lagrange multiplier is zero unless the ith constraint is active at the optimum.

3) KKT optimality conditions: We now assume that the functions $f_0, \dots, f_m, h_1, \dots, h_p$ are differentiable (and therefore have open domains), but we make no assumptions yet about convexity.

Let $x^\star$ and $(\lambda^\star, \nu^\star)$ be any primal and dual optimal points with zero duality gap. Since $x^\star$ minimizes $L(x, \lambda^\star, \nu^\star)$ over x, it follows that its gradient must vanish at $x^\star$, *i.e.*,

$$\nabla f_0(x^\star) + \sum_{i=1}^m \lambda_i^\star \nabla f_i(x^\star) + \sum_{i=1}^p \nu_i^\star \nabla h_i(x^\star) = 0.$$

Thus we have

$$\begin{aligned}
f_i(x^\star) &\le 0, \quad i = 1, \dots, m \\
h_i(x^\star) &= 0, \quad i = 1, \dots, p \\
\lambda_i^\star &\succeq 0, \quad i = 1, \dots, m \\
\lambda_i^\star f_i(x^\star) &= 0, \quad i = 1, \dots, m \\
\nabla f_0(x^\star) + \sum_{i=1}^m \lambda_i^\star \nabla f_i(x^\star) + \sum_{i=1}^p \nu_i^\star \nabla h_i(x^\star) &= 0,
\end{aligned} \tag{13}$$

which are called the *Karush-Kuhn-Tucker* (KKT) conditions.

To summarize, for *any* optimization problem with differentiable objective and constraint functions for which strong duality obtains, any pair of primal and dual optimal points must satisfy the KKT conditions (13).

Now when the primal problem is convex, the KKT conditions are also sufficient for the points to be primal and dual optimal. In other words, if f_i are convex and h_i are affine, and $\tilde{x}, \tilde{\lambda}, \tilde{\nu}$ are any points that satisfy the KKT conditions

$$\begin{aligned}
f_i(\tilde{x}) &\le 0, \quad i = 1, \dots, m \\
h_i(\tilde{x}) &= 0, \quad i = 1, \dots, p \\
\tilde{\lambda}_i &\ge 0, \quad i = 1, \dots, m \\
\tilde{\lambda}_i f_i(\tilde{x}) &= 0, \quad i = 1, \dots, m \\
\nabla f_0(\tilde{x}) + \sum_{i=1}^m \tilde{\lambda}_i \nabla f_i(\tilde{x}) + \sum_{i=1}^p \tilde{\nu}_i \nabla h_i(\tilde{x}) &= 0,
\end{aligned}$$

then $\tilde{x}$ and $(\tilde{\lambda}, \tilde{\nu})$ are primal and dual optimal, with zero duality gap.

If a convex optimization problem with differentiable objective and constraint functions satisfies Slater's condition, then the KKT conditions provide necessary and sufficient conditions for optimality: Slater's condition implies that the optimal duality gap is zero and the dual optimum is attained, so x is optimal if and only if there are (λ, ν) that, together with x, satisfy the KKT conditions.

The KKT conditions play an important role in optimization. In a few special cases it is possible to solve the KKT conditions (and therefore, the optimization problem) analytically. More generally, many algorithms for convex optimization, such as the barrier method in section III, are conceived as, or can be interpreted as, methods for solving the KKT conditions.

Example *Equality constrained convex quadratic minimization.* We consider the quadratic programming (QP) problem

$$\begin{aligned}
\text{minimize} \quad & (1/2)x^T P x + q^T x + r \\
\text{subject to} \quad & Ax = b,
\end{aligned} \tag{14}$$

where $P \in \mathbf{S}_+^n$. The KKT conditions for this problem are

$$Ax^\star = b, \qquad Px^\star + q + A^T \nu^\star = 0,$$

which we can write as

$$\begin{bmatrix} P & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x^\star \\ \nu^\star \end{bmatrix} = \begin{bmatrix} -q \\ b \end{bmatrix}.$$

Solving this set of $m+n$ equations in the $m+n$ variables $x^\star, \nu^\star$ gives the optimal primal and dual variables for (14). This set of equations is called the *KKT system* for the equality constrained quadratic optimization problem (14). The coefficient matrix is called the *KKT matrix*.

When the KKT matrix is nonsingular, there is a unique optimal primal-dual pair $(x^\star, \nu^\star)$. If the KKT matrix is singular, but the KKT system is solvable, any solution yields an optimal pair $(x^\star, \nu^\star)$. If the KKT system is not solvable, the quadratic optimization problem is unbounded below. Indeed, in this case there exist $v \in \mathbf{R}^n$ and $w \in \mathbf{R}^p$ such that

$$Pv + A^T w = 0, \qquad Av = 0, \qquad -q^T v + b^T w > 0.$$

Let $\hat{x}$ be any feasible point. The point $x = \hat{x} + tv$ is feasible for all t and

$$\begin{aligned}
f(\hat{x} + tv) &= f(\hat{x}) + t(v^T P \hat{x} + q^T v) + (1/2)t^2 v^T P v \\
&= f(\hat{x}) + t(-\hat{x}^T A^T w + q^T v) - (1/2)t^2 w^T A v \\
&= f(\hat{x}) + t(-b^T w + q^T v),
\end{aligned}$$

which decreases without bound as $t \to \infty$.

---

4) Solving the primal problem via the dual: We mentioned at the beginning of §II-C.3 that if strong duality holds and a dual optimal solution $(\lambda^\star, \nu^\star)$ exists, then any primal optimal point is also a minimizer of $L(x, \lambda^\star, \nu^\star)$. This fact sometimes allows us to compute a primal optimal solution from a dual optimal solution.

More precisely, suppose we have strong duality and an optimal $(\lambda^\star, \nu^\star)$ is known. Suppose that the minimizer of $L(x, \lambda^\star, \nu^\star)$, *i.e.*, the solution of

$$\text{minimize} \quad f_0(x) + \sum_{i=1}^m \lambda_i^\star f_i(x) + \sum_{i=1}^p \nu_i^\star h_i(x), \tag{15}$$

is unique. (For a convex problem this occurs, for example, if $L(x, \lambda^\star, \nu^\star)$ is a strictly convex function of x.) Then if the solution of (15) is primal feasible, it must be primal optimal; if it is not primal feasible, then no primal optimal point can exist, *i.e.*, we can conclude that the primal optimum is not attained. This observation is interesting when the dual problem is easier to solve than the primal problem, for example, because it can be solved analytically, or has some special structure that can be exploited. The following simple example shows how this can be used to decompose a large problem into small managable subproblems.

Example *Minimizing a separable function subject to an equality constraint.* We consider a problem which would made up of a large number (n) of independent minimization problems, were it not for a few coupling constraints (one in this case):

$$
\begin{aligned}
\text{minimize} \quad & f_0(x) = \sum_{i=1}^n f_i(x_i) \\
\text{subject to} \quad & a^T x = b,
\end{aligned}
$$

where $a \in \mathbf{R}^n$, $b \in \mathbf{R}$, and $f_i : \mathbf{R} \to \mathbf{R}$ are differentiable and *strictly* convex. The objective function is called *separable* since it is a sum of functions of the individual variables $x_1, \ldots, x_n$. We assume that the domain of f_0 intersects the constraint set, *i.e.*, there exists a point $x_0 \in \mathbf{dom}\, f$ with $a^T x_0 = b$. This implies the problem has a unique optimal point $x^\star$.
The Lagrangian is

$$
L(x, \nu) = \sum_{i=1}^n f_i(x_i) + \nu(a^T x - b) = -b\nu + \sum_{i=1}^n (f_i(x_i) + \nu a_i x_i),
$$

which is also separable, so the dual function is

$$
\begin{aligned}
g(\nu) &= -b\nu + \inf_x \left(\sum_{i=1}^n (f_i(x_i) + \nu a_i x_i) \right) \\
&= -b\nu + \sum_{i=1}^n \inf_{x_i} (f_i(x_i) + \nu a_i x_i) \\
&= -b\nu - \sum_{i=1}^n f_i^*(-\nu a_i).
\end{aligned}
$$

where $f_i^*(-\nu a_i) \triangleq \inf_{x_i}(f_i(x_i) + \nu a_i x_i)$. The dual problem is thus

$$
\text{maximize} \quad -b\nu - \sum_{i=1}^n f_i^*(-\nu a_i),
$$

with (scalar) variable $\nu \in \mathbf{R}$.
Assume that f_i^*'s are easy to evaluate and therefore that the dual problem is easy to solve. Suppose that we have found an optimal dual variable $\nu^\star$. (There are several simple methods for solving a convex problem with one scalar variable, such as the bisection method.) Since each f_i is strictly convex, the function $L(x, \nu^\star)$ is strictly convex in x, and so has a unique minimizer $\tilde{x}$. But we also know that $x^\star$ minimizes $L(x, \nu^\star)$, so we must have $\tilde{x} = x^\star$. We can recover $x^\star$ from $\nabla_x L(x, \nu^\star) = 0$, *i.e.*, by solving the equations $f_i'(x_i^\star) = -\nu^\star a_i$.

D. Perturbation and sensitivity analysis

When strong duality obtains, the optimal dual variables give very useful information about the sensitivity of the optimal value with respect to perturbations of the constraints.

1) The perturbed problem: We consider the following perturbed version of the original optimization problem (1):

$$
\begin{aligned}
\text{minimize} \quad & f_0(x) \\
\text{subject to} \quad & f_i(x) \le u_i, \quad i = 1, \ldots, m \\
& h_i(x) = v_i, \quad i = 1, \ldots, p,
\end{aligned}
\tag{16}
$$

with variable $x \in \mathbf{R}^n$. This problem coincides with the original problem (1) when $u = 0$, $v = 0$. When u_i is positive it means that we have relaxed the ith inequality constraint; when u_i is negative, it means that we have tightened the constraint. Thus the perturbed problem (16) results from the original problem (1) by tightening or relaxing each inequality constraint by u_i, and changing the righthand side of the equality constraints by v_i.

We define $p^\star(u, v)$ as the optimal value of the perturbed problem (16):

$$
p^\star(u, v) = \inf\{f_0(x) \mid \exists x \in \mathcal{D}, \ f_i(x) \le u_i, \ i = 1, \ldots, m, \\
h_i(x) = v_i, \ i = 1, \ldots, p\}.
$$

We can have $p^\star(u, v) = \infty$, which corresponds to perturbations of the constraints that result in infeasibility. Note that $p^\star(0, 0) = p^\star$, the optimal value of the unperturbed problem (1). (We hope this slight abuse of notation will cause no confusion.) Roughly speaking, the function $p^\star : \mathbf{R}^m \times \mathbf{R}^p$ gives the optimal value of the problem as a function of perturbations to the righthand sides of the constraints.

When the original problem is convex, the function $p^\star$ turns out to be a convex function of u and v [8].

2) A global inequality: Now we assume that strong duality holds, and that the dual optimum is attained. (This is the case if the original problem is convex, and Slater's condition is satisfied). Let $(\lambda^\star, \nu^\star)$ be optimal for the dual (4) of the unperturbed problem. Then for all u and v we have [8]

$$
p^\star(u, v) \ge p^\star(0, 0) - \lambda^{\star T} u - \nu^{\star T} v.
\tag{17}
$$

Various sensitivity interpretations of the optimal Lagrange variables follow directly from the inequality (17) [8]. For example, if $\lambda_i^\star$ is large and we tighten the ith constraint (*i.e.*, choose $u_i < 0$), then the optimal value $p^\star(u, v)$ is guaranteed to increase greatly; and if $\lambda_i^\star$ is small, and we loosen the ith constraint ($u_i > 0$), then the optimal value $p^\star(u, v)$ will not decrease too much.

The inequality (17), and the conclusions listed above, give a *lower bound* on the perturbed optimal value, but no upper bound. For this reason the results are *not* symmetric with respect to loosening or tightening a constraint. For example, suppose that $\lambda_i^\star$ is large, and we loosen the ith constraint a bit (*i.e.*, take u_i small and positive). In this case the inequality (17) is not useful; it does not, for example, imply that the optimal value will decrease considerably.

3) Local sensitivity analysis: Suppose now that $p^\star(u, v)$ is differentiable at $u = 0$, $v = 0$. Then, provided strong duality holds, the optimal dual variables $\lambda^\star$, $\nu^\star$ are related to the gradient of $p^\star$ at $u = 0$, $v = 0$ [8]:

$$
\lambda_i^\star = -\frac{\partial p^\star(0, 0)}{\partial u_i}, \qquad \nu_i^\star = -\frac{\partial p^\star(0, 0)}{\partial v_i}.
\tag{18}
$$

Thus, when $p^\star(u, v)$ is differentiable at $u = 0$, $v = 0$, and strong duality holds, the optimal Lagrange multipliers are exactly the local sensitivities of the optimal value with respect to constraint perturbations.

III. INTERIOR POINT ALGORITHMS

In the late '80's and 90's there was a breakthrough: Karmarkar's polynomial complexity interior point method for Linear Programming was extended to a much wider class of convex optimization problems such as QP's, SOCP's, SDP's, etc [21], [27], [32], [33]. In this section, we will summarize the barrier method as a representative technique from the class of interior point methods. The barrier method is simple to implement and shows good performance in general. We will focus mainly on the key algorithmic components for practical implementation of interior point methods, along with intuition as to why these methods work. These key components are: backtracking linesearch, the Newton method for equality constrained minimization, and the logarithmic barrier function.

We will also show how duality plays a critical and practical role in these algorithms. In particular, using duality, interior point methods provide exact stopping criteria for terminating the search. The user supplies their desired tolerance, and when the search terminates, the returned decision vector is guaranteed to be within the specified tolerance of the optimum. This is in contrast to other methods which terminate simply when the rate of progress becomes slow, but without any guarantees on the optimality of the returned decision vector.

A. Descent methods and linesearch

Suppose we want to solve the unconstrained optimization problem

$$\text{minimize} \quad f(x) \tag{19}$$

where $f : \mathbf{R}^n \to \mathbf{R}$ is convex and twice continuously differentiable (which implies that $\operatorname{\mathbf{dom}} f$ is open). We will assume that the problem is solvable, *i.e.*, there exists an optimal point $x^\star$. (More precisely, the assumptions later in the section will imply that $x^\star$ exists and is unique.) We denote the optimal value, $\inf_x f(x) = f(x^\star)$, as $p^\star$.

Since f is differentiable and convex, a necessary and sufficient condition for a point $x^\star$ to be optimal is

$$\nabla f(x^\star) = 0. \tag{20}$$

Thus, solving the unconstrained minimization problem (19) is the same as finding a solution of (20), which is a set of n equations in the n variables $x_1, \ldots, x_n$. In a few special cases, we can find a solution to the problem (19) by analytically solving the optimality equation (20), but usually the problem must be solved by an iterative algorithm. By this we mean an algorithm that computes a sequence of points $x^{(0)}, x^{(1)}, \ldots \in \operatorname{\mathbf{dom}} f$ with $f(x^{(k)}) \to p^\star$ as $k \to \infty$. Such a sequence of points is called a *minimizing sequence* for the problem (19). The algorithm is terminated when $f(x^{(k)}) - p^\star \le \epsilon$, where $\epsilon > 0$ is some specified tolerance.

1) Descent methods: The algorithms described in this section produce a minimizing sequence $x^{(k)}$, $k = 1, \ldots,$ where

$$x^{(k+1)} = x^{(k)} + t^{(k)} \Delta x^{(k)}$$

and $t^{(k)} > 0$ (except when $x^{(k)}$ is optimal). Here the concatenated symbols Δ and x that form Δx are to be read as a single entity, a vector in $\mathbf{R}^n$ called the *step* or *search direction* (even though it need not have unit norm), and $k = 0, 1, \ldots$ denotes the iteration number. The scalar $t^{(k)} \ge 0$ is called the *step size* or *step length* at iteration k (even though it is not equal to $\|x^{(k+1)} - x^{(k)}\|$ unless $\|\Delta x^{(k)}\| = 1$).

The methods we present here are *descent methods*, which means that

$$f(x^{(k+1)}) < f(x^{(k)}),$$

except when $x^{(k)}$ is optimal. This implies that for all k we have $x^{(k)} \in S$, the initial sublevel set, and in particular we have $x^{(k)} \in \operatorname{\mathbf{dom}} f$. From convexity we know that $\nabla f(x^{(k)})^T (y - x^{(k)}) \ge 0$ implies $f(y) \ge f(x^{(k)})$, so the search direction in a descent method must satisfy

$$\nabla f(x^{(k)})^T \Delta x^{(k)} < 0,$$

i.e., it must make an acute angle with the negative gradient. We call such a direction a *descent direction* (for f, at $x^{(k)}$).

The outline of a general descent method is as follows. It alternates between two steps: determining a descent direction Δx, and the selection of a step size t.

Algorithm *General descent method.*
given a starting point $x \in \operatorname{\mathbf{dom}} f$.
repeat
    1. Determine a descent direction Δx.
    2. *Line search.* Choose a step size $t > 0$.
    3. *Update.* $x := x + t\Delta x$.
until stopping criterion is satisfied.

The second step is called the *line search* since selection of the step size t determines where along the line $\{x + t\Delta x \mid t \in \mathbf{R}_+\}$ the next iterate will be.

2) Backtracking line search: Most line searches used in practice are *inexact*: the step length is chosen to approximately minimize f along the ray $\{x + t\Delta x \mid t \ge 0\}$, or even to just reduce f 'enough'. Many inexact line search methods have been proposed. One inexact line search method that is very simple and quite effective is called *backtracking* line search. It depends on two constants α, β with $0 < \alpha < 0.5$, $0 < \beta < 1$.

Algorithm *Backtracking line search.*
given a descent direction Δx for f at $x \in \operatorname{\mathbf{dom}} f$,
    $\alpha \in (0, 0.5)$, $\beta \in (0, 1)$.
$t := 1$.
while $f(x + t\Delta x) > f(x) + \alpha t \nabla f(x)^T \Delta x$, $t := \beta t$.

The line search is called backtracking because it starts with unit step size and then reduces it by the factor β until the stopping condition $f(x + t\Delta x) \le f(x) +$

$\alpha t \nabla f(x)^T \Delta x$ holds. Since Δx is a descent direction, we have $\nabla f(x)^T \Delta x < 0$, so for small enough t we have

$$f(x + t\Delta x) \approx f(x) + t\nabla f(x)^T \Delta x < f(x) + \alpha t \nabla f(x)^T \Delta x,$$

which shows that the backtracking line search eventually terminates. The constant α can be interpreted as the fraction of the decrease in f predicted by linear extrapolation that we will accept. See [8] for how to choose the paramters t, α, and β.

There is still the question of exactly how to compute the descent direction Δx. We will explain how this can be done next, using the general Newton method for equality constrained minimization.

B. Newton method for equality constrained minimization

In this section we describe the Newton method for solving a convex optimization problem with equality constraints,

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & Ax = b, \end{array} \tag{21}$$

where $f : \mathbf{R}^n \to \mathbf{R}$ is convex and twice continuously differentiable, and $A \in \mathbf{R}^{p \times n}$ with $\mathbf{rank}\, A = p < n$. The assumptions on A mean that there are fewer equality constraints than variables, and that the equality constraints are independent. We will assume that an optimal solution $x^\star$ exists, and use $p^\star$ to denote the optimal value, $p^\star = \inf\{f(x) \mid Ax = b\} = f(x^\star)$.

Any equality constrained minimization problem can be reduced to an equivalent unconstrained problem by eliminating the equality constraints and then solving an unconstrained problem. However, we will focus on an extension of Newton's method that directly handles equality rather than methods that eliminate the inequalities. One reason is that problem structure, such as sparsity, is often destroyed by elimination (or forming the dual); in contrast, a method that directly handles equality constraints can exploit the problem structure. Another reason is conceptual: methods that directly handle equality constraints can be thought of as methods for directly solving the optimality conditions (22), shown below.

Recall (from §II-C.3) that a point $x^\star \in \mathbf{dom}\, f$ is optimal for (21) if and only if there is a $\nu^\star \in \mathbf{R}^p$ such that

$$Ax^\star = b, \qquad \nabla f(x^\star) + A^T \nu^\star = 0. \tag{22}$$

Solving the equality constrained optimization problem (21) is therefore equivalent to finding a solution of the KKT equations (22), which is a set of $n + p$ equations in the $n + p$ variables $x^\star, \nu^\star$. As with unconstrained optimization, there are a few problems for which we can solve these optimality conditions analytically. The most important special case is when f is quadratic, as we saw in an earlier example.

1) The Newton Step: The barrier method, and in fact most interior point methods, use Newton's method to compute the descent direction. Of all known techniques for computing descent directions, Newton's method exhibits among the fastest convergence rates. It is also one of the most expensive computationally. However, when used in a carefully designed interior point method, its provable speed and performace more than compensate for its computational cost [8].

The superior performance of Newton's method is due to the fact that it explicitly uses second derivative information from the Hessian of the objective in its computation of the descent direction. In particular, it chooses the direction, along the set defined by the equality constraints, which minimizes a local quadratic approximation of the objective at the last iteration. The Newton step Δx_{nt} is what must be added to x to solve the problem when the quadratic approximation is used in place of f in (21). The Newton step Δx_{nt} for the equality constrained optimization problem (21) is characterized by

$$\begin{bmatrix} \nabla^2 f(x) & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x_{\mathrm{nt}} \\ w \end{bmatrix} = \begin{bmatrix} -\nabla f(x) \\ 0 \end{bmatrix}, \tag{23}$$

where w is the associated optimal dual variable for the quadratic problem. The Newton step is defined only at points for which the KKT matrix is nonsingular. This condition will be ensured in any proper implementation of the barrier method.

As in Newton's method for unconstrained problems, we observe that when the objective f is exactly quadratic, the Newton update $x + \Delta x_{\mathrm{nt}}$ exactly solves the equality constrained minimization problem, and in this case the vector w is the optimal dual variable for the original problem. This suggests, as in the unconstrained case, that when f is nearly quadratic, $x + \Delta x_{\mathrm{nt}}$ should be a very good estimate of the solution $x^\star$, and w should be a good estimate of the optimal dual variable $\nu^\star$. Also note that in the special case of unconstrained minimization, Newton's method reduces to solving:

$$\Delta x_{\mathrm{nt}} = -(\nabla^2 f(x))^{-1} \nabla f(x)$$

2) Feasible directions and the Newton decrement: We define the Newton decrement for the equality constrained problem as

$$\lambda(x) = (\Delta x_{\mathrm{nt}}^T \nabla^2 f(x) \Delta x_{\mathrm{nt}})^{1/2}. \tag{24}$$

It can be shown [8] that $\lambda(x)^2/2$ gives an estimate of $f(x) - p^\star$, based on the quadratic model at x, and also that $\lambda(x)$ (or a multiple of $\lambda(x)^2$) serves as the basis of a good stopping criterion.

We say that $v \in \mathbf{R}^n$ is a *feasible direction* if $Av = 0$. Now suppose that $Ax = b$. In this case, every point of the form $x + tv$ is also feasible, *i.e.*, $A(x + tv) = b$. We say that v is a *descent direction* for f at x, if for small $t > 0$, $f(x + tv) < f(x)$.

The Newton step is always a feasible descent direction (except when x is optimal, in which case $\Delta x_{\mathrm{nt}} = 0$). Indeed, the second set of equations that define Δx_{nt} are $A\Delta x_{\mathrm{nt}} = 0$, which shows it is a feasible direction; that it is a descent direction follows from the fact that the directional derivative of f along Δx_{nt} is negative - in fact it is exactly $-\lambda(x)^2$ [8].

3) Newton's method with equality constraints: Thus we arrive at the desired Newton method with equality constraints:

Algorithm *Newton's method for equality constrained minimization.*

given starting point $x \in \mathbf{dom}\, f$ with $Ax = b$, tolerance $\epsilon > 0$.

repeat
1. Compute the Newton step and decrement Δx_{nt}, $\lambda(x)$.
2. *Stopping criterion.* **quit** if $\lambda^2/2 \le \epsilon$.
3. *Line search.* Choose step size t by backtracking line search.
4. *Update.* $x := x + t\Delta x_{\mathrm{nt}}$.

The method is called a *feasible descent method*, since all the iterates are feasible, with $f(x^{(k+1)}) < f(x^{(k)})$ (unless $x^{(k)}$ is optimal). Newton's method requires that the KKT matrix be invertible at each x.

C. Barrier method for inequality constrained minimization

We now move on to *interior-point methods* for solving convex optimization problems that include inequality constraints,

$$
\begin{aligned}
\text{minimize} \quad & f_0(x) \\
\text{subject to} \quad & f_i(x) \le 0, \quad i = 1, \ldots, m \\
& Ax = b,
\end{aligned}
\tag{25}
$$

where $f_0, \ldots, f_m : \mathbf{R}^n \to \mathbf{R}$ are convex and twice continuously differentiable, and $A \in \mathbf{R}^{p \times n}$ with $\mathbf{rank}\, A = p < n$. We assume that the problem is solvable, *i.e.*, an optimal $x^\star$ exists. We denote the optimal value $f_0(x^\star)$ as $p^\star$.

We also assume that the problem is strictly feasible, *i.e.*, there exists $x \in \mathcal{D}$ that satisfies $Ax = b$ and $f_i(x) < 0$ for $i = 1, \ldots, m$. This means that Slater's constraint qualification holds, so there exist dual optimal $\lambda^\star \in \mathbf{R}^m$, $\nu^\star \in \mathbf{R}^p$, which together with $x^\star$ satisfy the KKT conditions

$$
\begin{aligned}
Ax^\star = b, \quad f_i(x^\star) &\le 0, \quad i = 1, \ldots, m \\
\lambda_i^\star f_i(x^\star) &= 0, \quad i = 1, \ldots, m \\
\lambda^\star &\succeq 0 \\
\nabla f_0(x^\star) + \sum_{i=1}^m \lambda_i^\star \nabla f_i(x^\star) &+ A^T \nu^\star = 0.
\end{aligned}
\tag{26}
$$

Interior-point methods solve the problem (25) (or the KKT conditions (26)) by applying Newton's method to a sequence of equality constrained problems, or to a sequence of modified versions of the KKT conditions. We will concentrate on a particular interior-point algorithm, the *barrier method*.

Many problems are already in the form (25), and satisfy the assumption that the objective and constraint functions are twice differentiable. Obvious examples are LPs, QPs, QCQPs, and GPs in convex form. Many other problems do not have the required form (25), with twice differentiable objective and constraint functions, but can be reformulated in the required form. We have already seen many examples of this, such as the transformation of an unconstrained convex piecewise-linear minimization problem. Other convex optimization problems, such as SOCPs and SDPs, are not readily recast in the required form, but can be handled by extensions of interior-point methods to problems with generalized inequalities, see [8].

1) Logarithmic barrier function: Now suppose, for the moment, that our goal is to approximately formulate the inequality constrained problem (25) as an equality constrained problem to which Newton's method can be applied. Consider the approximation

$$
\begin{aligned}
\text{minimize} \quad & f_0(x) + \sum_{i=1}^m -(1/t) \log(-f_i(x)) \\
\text{subject to} \quad & Ax = b.
\end{aligned}
\tag{27}
$$

The objective here is convex, since $-(1/t)\log(-u)$ is convex and increasing in u, and differentiable. Assuming an appropriate closedness condition holds, Newton's method can be used to solve it.

The function

$$
\phi(x) = -\sum_{i=1}^m \log(-f_i(x)),
\tag{28}
$$

with $\mathbf{dom}\, \phi = \{x \in \mathbf{R}^n \mid f_i(x) < 0, \ i = 1, \ldots, m\}$, is called the *logarithmic barrier* or *log barrier* for the problem (25). Its domain is the set of points that satisfy the inequality constraints of (25) strictly. No matter what value the positive parameter t has, the logarithmic barrier grows without bound if $f_i(x) \to 0$, for any i.

Of course, the problem (27) is only an approximation of the original problem (25), so one question that arises immediately is how well a solution of (27) approximates a solution of the original problem (25). Intuition suggests, and we will soon confirm, that the quality of the approximation improves as the parameter t grows.

On the other hand, when the parameter t is large, the function $f_0 + (1/t)\phi$ is difficult to minimize by Newton's method, since its Hessian varies rapidly near the boundary of the feasible set. We will see that this problem can be circumvented by solving a *sequence* of problems of the form (27), increasing the parameter t (and therefore the accuracy of the approximation) at each step, and starting each Newton minimization at the solution of the problem for the previous value of t.

For future reference, we note that the gradient and Hessian of the logarithmic barrier function ϕ are given by

$$
\nabla \phi(x) = \sum_{i=1}^m \frac{1}{-f_i(x)} \nabla f_i(x),
$$

$$
\nabla^2 \phi(x) = \sum_{i=1}^m \frac{1}{f_i(x)^2} \nabla f_i(x) \nabla f_i(x)^T + \sum_{i=1}^m \frac{1}{-f_i(x)} \nabla^2 f_i(x)
$$

(these expressions follow from the chain rule and some tedious manipulations, see [8]).

2) Central path: We now consider in more detail the minimization problem (27). It will simplify notation later on if we multiply the objective by t, and consider the equivalent problem

$$
\begin{aligned}
\text{minimize} \quad & tf_0(x) + \phi(x) \\
\text{subject to} \quad & Ax = b,
\end{aligned}
\tag{29}
$$

which has the same minimizers. We assume for now that the problem (29) can be solved via Newton's method, and, in particular, that it has a unique solution for each $t > 0$.

For $t > 0$ we define $x^\star(t)$ as the solution of (29). The *central path* associated with problem (25) is defined as the set of points $x^\star(t)$, $t > 0$, which we call the *central points*.

Points on the central path are characterized by the following necessary and sufficient conditions: $x^\star(t)$ is strictly feasible, *i.e.*, satisfies

$$Ax^\star(t) = b, \qquad f_i(x^\star(t)) < 0, \quad i = 1, \dots, m,$$

and there exists a $\hat\nu \in \mathbf{R}^p$ such that

$$
\begin{aligned}
0 &= t\nabla f_0(x^\star(t)) + \nabla\phi(x^\star(t)) + A^T\hat\nu \\
&= t\nabla f_0(x^\star(t)) + \sum_{i=1}^m \frac{1}{-f_i(x^\star(t))}\nabla f_i(x^\star(t)) \\
&\quad + A^T\hat\nu
\end{aligned}
\tag{30}
$$

holds.

Example *Inequality form linear programming.* The logarithmic barrier function for an LP in inequality form,

$$
\begin{array}{ll}
\text{minimize} & c^T x \\
\text{subject to} & Ax \preceq b,
\end{array}
\tag{31}
$$

is given by

$$\phi(x) = -\sum_{i=1}^m \log(b_i - a_i^T x), \qquad \mathbf{dom}\,\phi = \{x \mid Ax \prec b\},$$

where $a_1^T, \dots, a_m^T$ are the rows of A. The gradient and Hessian of the barrier function are

$$\nabla\phi(x) = \sum_{i=1}^m \frac{1}{b_i - a_i^T x}a_i, \quad \nabla^2\phi(x) = \sum_{i=1}^m \frac{1}{(b_i - a_i^T x)^2}a_i a_i^T,$$

or, more compactly,

$$\nabla\phi(x) = A^T d, \; \nabla^2\phi(x) = A^T\,\mathbf{diag}(d)^2 A,$$

where the elements of $d \in \mathbf{R}^m$ are given by $d_i = 1/(b_i - a_i^T x)$. Since x is strictly feasible, we have $d \succ 0$, so the Hessian of ϕ is nonsingular if and only if A has rank n.
The centrality condition (30) is

$$tc + \sum_{i=1}^m \frac{1}{b_i - a_i^T x}a_i = tc + A^T d = 0. \tag{32}$$

We can give a simple geometric interpretation of the centrality condition. At a point $x^\star(t)$ on the central path the gradient $\nabla\phi(x^\star(t))$, which is normal to the level set of ϕ through $x^\star(t)$, must be parallel to $-c$. In other words, the hyperplane $c^T x = c^T x^\star(t)$ is tangent to the level set of ϕ through $x^\star(t)$.

---

3) Dual points from central path: From (30) we can derive an important property of the central path: Every central point yields a dual feasible point, and hence a lower bound on the optimal value $p^\star$. More specifically, define

$$\lambda_i^\star(t) = -\frac{1}{tf_i(x^\star(t))}, \quad i = 1, \dots, m, \qquad \nu^\star(t) = \hat\nu/t. \tag{33}$$

We claim that the pair $\lambda^\star(t)$, $\nu^\star(t)$ is dual feasible.

First, it is clear that $\lambda^\star(t) \succ 0$ because $f_i(x^\star(t)) < 0$, $i = 1, \dots, m$. By expressing the optimality conditions (30) as

$$\nabla f_0(x^\star(t)) + \sum_{i=1}^m \lambda_i^\star(t)\nabla f_i(x^\star(t)) + A^T\nu^\star(t) = 0,$$

we see that $x^\star(t)$ minimizes the Lagrangian

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^m \lambda_i f_i(x) + \nu^T(Ax - b),$$

for $\lambda = \lambda^\star(t)$ and $\nu = \nu^\star(t)$, which means that $\lambda^\star(t)$, $\nu^\star(t)$ is a dual feasible pair. Therefore the dual function $g(\lambda^\star(t), \nu^\star(t))$ is finite, and

$$
\begin{aligned}
g(\lambda^\star(t), \nu^\star(t)) &= f_0(x^\star(t)) + \sum_{i=1}^m \lambda_i^\star(t)f_i(x^\star(t)) \\
&\quad + \nu^\star(t)^T(Ax^\star(t) - b) \\
&= f_0(x^\star(t)) - m/t.
\end{aligned}
$$

In particular, the duality gap associated with $x^\star(t)$ and the dual feasible pair $\lambda^\star(t)$, $\nu^\star(t)$ is simply m/t. As an important consequence, we have

$$f_0(x^\star(t)) - p^\star \le m/t,$$

i.e., $x^\star(t)$ is no more than m/t suboptimal. This confirms the intuitive idea that $x^\star(t)$ converges to an optimal point as $t \to \infty$.

Example *Inequality form linear programming.* The dual of the inequality form LP (31) is

$$
\begin{array}{ll}
\text{maximize} & -b^T\lambda \\
\text{subject to} & A^T\lambda + c = 0 \\
& \lambda \succeq 0.
\end{array}
$$

From the optimality conditions (32), it is clear that

$$\lambda_i^\star(t) = \frac{1}{t(b_i - a_i^T x^\star(t))}, \quad i = 1, \dots, m,$$

is dual feasible, with dual objective value

$$-b^T\lambda^\star(t) = c^T x^\star(t) + (Ax^\star(t) - b)^T\lambda^\star(t) = c^T x^\star(t) - m/t.$$

---

4) The barrier method: We have seen that the point $x^\star(t)$ is m/t-suboptimal, and that a certificate of this accuracy is provided by the dual feasible pair $\lambda^\star(t)$, $\nu^\star(t)$. This suggests a very straightforward method for solving the original problem (25) with a guaranteed specified accuracy ϵ: We simply take $t = m/\epsilon$ and solve the equality constrained problem

$$
\begin{array}{ll}
\text{minimize} & (m/\epsilon)f_0(x) + \phi(x) \\
\text{subject to} & Ax = b
\end{array}
$$

using Newton's method. This method could be called the *unconstrained minimization method*, since it allows us to solve the inequality constrained problem (25) to a guaranteed accuracy by solving an unconstrained, or linearly constrained, problem. Although this method can work well for small problems, good starting points, and moderate accuracy (*i.e.*, ϵ not too small), it does not work well in other cases because the Hessian tends to vary too rapidly near the "edges" of the barrier, which is where optima often tend to lie. As a result it is rarely, if ever, used.

A simple extension of the unconstrained minimization method does work extremely well. When the method was first proposed in the 1960s, it was called the *sequential unconstrained minimization technique* (SUMT). Somewhat recently [27], it has been shown to have polynomial time

convergence for a wide class of convex optimization problems [27]. Today the method is usually called the *barrier method* or *path-following method*.

The barrier method is based on solving a sequence of unconstrained (or linearly constrained) minimization problems, using the last point found as the starting point for the next unconstrained minimization problem. In other words, we compute $x^\star(t)$ for a sequence of increasing values of t, until $t \geq m/\epsilon$, which guarantees that we have an ϵ-suboptimal solution of the original problem.

Thus we finally arrive at our desired:

Algorithm *Barrier method.*
given strictly feasible x, $t := t^{(0)} > 0$, $\mu > 1$, tolerance $\epsilon > 0$.
repeat
    1. *Centering step.* Compute $x^\star(t)$ by minimizing $tf_0 + \phi$, subject to $Ax = b$, starting at x.
    2. *Update.* $x := x^\star(t)$.
    3. *Stopping criterion.* **quit** if $m/t < \epsilon$.
    4. *Increase t.* $t := \mu t$.

At each iteration (except the first one) we compute the central point $x^\star(t)$ starting from the previously computed central point, and then increase t by a factor $\mu > 1$. The algorithm can also return $\lambda = \lambda^\star(t)$, and $\nu = \nu^\star(t)$, a dual ϵ-suboptimal point, or certificate for x.

Each execution of step 1 is called a *centering step* (since a central point is being computed) or an *outer iteration*, and to the first centering step (the computation of $x^\star(t^{(0)})$) as the *initial centering step*. (Thus the simple algorithm with $t^{(0)} = m/\epsilon$ consists of only the initial centering step.)

Newton's method is used in step 1 for linearly constrained minimization. In this case, the Newton step Δx_{nt}, and associated dual variable are given by the linear equations

$$\begin{bmatrix} H & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \Delta x_{\mathrm{nt}} \\ \nu_{\mathrm{nt}} \end{bmatrix} = \begin{bmatrix} -g \\ 0 \end{bmatrix}. \qquad (34)$$

where $H = t\nabla^2 f_0(x) + \nabla^2 \phi(x)$ and $g = t\nabla f_0(x) + \nabla \phi(x)$. The Newton iterations or steps executed during the centering step are called *inner iterations*. At each inner step, we have a primal feasible point; we have a dual feasible point, however, only at the end of each outer (centering) step.

5) Feasibility and phase I methods: The barrier method requires a strictly feasible starting point $x^{(0)}$. When such a point is not known, the barrier method is preceded by a preliminary stage, called *phase I*, in which a strictly feasible point is computed (or the constraints are found to be infeasible). The strictly feasible point found during phase I is then used as the starting point for the barrier method, which is called the *phase II* stage. Thus in phase I, we must find a strictly feasible solution of the inequalities and equalities, or determine that none exists. To do this we form the following optimization problem:

$$\begin{array}{ll} \text{minimize} & s \\ \text{subject to} & f_i(x) \leq s, \quad i = 1, \ldots, m \\ & Ax = b \end{array} \qquad (35)$$

in the variables $x \in \mathbf{R}^n$, $s \in \mathbf{R}$. The variable s can be interpreted as a bound on the maximum infeasibility of the

inequalities; the goal is to drive the maximum infeasibility below zero.

This problem is always strictly feasible, since we can choose $x^{(0)}$ as starting point for x, and for s, we can choose any number larger than $\max_{i=1,\ldots,m} f_i(x^{(0)})$. We can therefore apply the barrier method to solve the problem (35), which is called the *phase I optimization problem* associated with the inequality and equality system (35).

IV. CONCLUSION

We hope that the reader has gained an appreciation for the elegance and practicality of duality and interior point methods. There is much that we haven't covered here, however we hope that the reader will be encouraged to learn more about convex optimization and its applications. The references and software tools cited in the introduction should provide good starting points, and new applications are constantly being discovered.

REFERENCES

[1] M. S. Bazaraa, H. D. Sherali, and C. M. Shetty. *Nonlinear Programming. Theory and Algorithms.* John Wiley & Sons, second edition, 1993.
[2] A. Ben-Tal and A. Nemirovski. *Lectures on Modern Convex Optimization. Analysis, Algorithms, and Engineering Applications.* Society for Industrial and Applied Mathematics, 2001.
[3] D. P. Bertsekas. *Nonlinear Programming.* Athena Scientific, 1999.
[4] D. Bertsimas and J. N. Tsitsiklis. *Introduction to Linear Optimization.* Athena Scientific, 1997.
[5] S. Boyd and C. Barratt. *Linear Controller Design: Limits of Performance.* Prentice-Hall, 1991.
[6] S. Boyd, P. Diaconis, and L. Xiao. Fastest mixing Markov chain on a graph. *SIAM Review*, 2003. Submitted.
[7] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan. *Linear Matrix Inequalities in System and Control Theory.* Society for Industrial and Applied Mathematics, 1994.
[8] S.P. Boyd and L. Vandenberghe. *Convex Optimization.* Cambridge University Press, 2004. Material available at www.stanford.edu/˜boyd.
[9] D. Colleran, C. Portmann, A. Hassibi, C. Crusius, S. Mohan, T. Lee S. Boyd, and M. Hershenson. Optimization of phase-locked loop circuits via geometric programming. In *Custom Integrated Circuits Conference (CICC)*, San Jose, California, September 2003.
[10] M. A. Dahleh and I. J. Diaz-Bobillo. *Control of Uncertain Systems: A Linear Programming Approach.* Prentice-Hall, 1995.
[11] J. W. Demmel. *Applied Numerical Linear Algebra.* Society for Industrial and Applied Mathematics, 1997.
[12] G. E. Dullerud and F. Paganini. *A Course in Robust Control Theory: A Convex Approach.* Springer, 2000.
[13] G. Golub and C. F. Van Loan. *Matrix Computations.* Johns Hopkins University Press, second edition, 1989.
[14] M. C. Grant, S. P. Boyd, and Y. Ye. *CVX: Matlab Software for Disciplined Convex Programming*, 2005. Available from www.stanford.edu/˜boyd/cvx/.
[15] M. Grötschel, L. Lovasz, and A. Schrijver. *Geometric Algorithms and Combinatorial Optimization.* Springer, 1988.
[16] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning. Data Mining, Inference, and Prediction.* Springer, 2001.
[17] M. del Mar Hershenson, S. P. Boyd, and T. H. Lee. Optimal design of a CMOS op-amp via geometric programming. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 20(1):1–21, 2001.
[18] H. Hindi. A tutorial on convex optimization. In *American Control Conference*, Boston, USA, June 2004.
[19] J.-B. Hiriart-Urruty and C. Lemaréchal. *Fundamentals of Convex Analysis.* Springer, 2001. Abridged version of *Convex Analysis and Minimization Algorithms* volumes 1 and 2.
[20] R. A. Horn and C. A. Johnson. *Matrix Analysis.* Cambridge University Press, 1985.

[21] N. Karmarkar. A new polynomial-time algorithm for linear programming. *Combinatorica*, 4(4):373–395, 1984.

[22] J. Löfberg. *YALMIP: a MATLAB toolbox for rapid prototyping of optimization problems. Version 3*, 2005. Available from `http://control.ee.ethz.ch/˜joloef/yalmip.msql` .

[23] D. G. Luenberger. *Optimization by Vector Space Methods*. John Wiley & Sons, 1969.

[24] D. G. Luenberger. *Linear and Nonlinear Programming*. Addison-Wesley, second edition, 1984.

[25] Z.-Q. Luo. Applications of convex optimization in signal processing and digital communication. *Mathematical Programming Series B*, 97:177–207, 2003.

[26] O. Mangasarian. *Nonlinear Programming*. Society for Industrial and Applied Mathematics, 1994. First published in 1969 by McGraw-Hill.

[27] Y. Nesterov and A. Nemirovskii. *Interior-Point Polynomial Methods in Convex Programming*. Society for Industrial and Applied Mathematics, 1994.

[28] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, 1970.

[29] G. Strang. *Linear Algebra and its Applications*. Academic Press, 1980.

[30] J. van Tiel. *Convex Analysis. An Introductory Text*. John Wiley & Sons, 1984.

[31] H. Wolkowicz, R. Saigal, and L. Vandenberghe, editors. *Handbook of Semidefinite Programming*. Kluwer Academic Publishers, 2000.

[32] S. J. Wright. *Primal-Dual Interior-Point Methods*. Society for Industrial and Applied Mathematics, 1997.

[33] Y. Ye. *Interior Point Algorithms. Theory and Analysis*. John Wiley & Sons, 1997.