Lecture 18: CS395T Numerical Optimization for Graphics and AI — Quadratic Programming

Qixing Huang
The University of Texas at Austin
huangqx@cs.utexas.edu

Disclaimer

This note is adapted from

- Section 16 and Section 18 of *Numerical Optimization* by Jorge Nocedal and Stephen J. Wright. Springer series in operations research and financial engineering. Springer, New York, NY, 2. ed. edition, (2006)

1 Quadratic Programming

An optimization problem with a quadratic objective function and linear constraints is called a quadratic program. Problems of this type are important in their own right, and they also arise as subproblems in methods for general constrained optimization, such as sequential quadratic programming, augmented Lagrangian methods, and interior-point methods.

The general quadratic program (QP) can be stated as

$$\min_{\boldsymbol{x}} \quad q(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T G\boldsymbol{x} + \boldsymbol{x}^T\boldsymbol{c} \tag{1}$$

$$\text{subject to} \quad \boldsymbol{a}_i^T\boldsymbol{x} = b_i, \quad i \in \mathcal{E}, \tag{2}$$

$$\boldsymbol{a}_i^T\boldsymbol{x} \geq b_i, \quad i \in \mathcal{I}, \tag{3}$$

where G is a symmetric $n \times n$ matrix, $\mathcal{E}$ and $\mathcal{I}$ are finite sets of indices, and $\boldsymbol{c}$, $\boldsymbol{x}$, and $\boldsymbol{a} - i, i \in \mathcal{E} \cup \mathcal{I}$, are vectors in $\mathbb{R}^n$. Quadratic programs can always be solved (or shown to be infeasible) in a finite amount of computation, but the effort required to find a solution depends strongly on the characteristics of the objective function and the number of inequality constraints. If the Hessian matrix G is positive semidefinite, we say that (3) is a convex QP, and in this case the problem is often similar in difficulty to a linear program. (Strictly convex QPs are those in which G is positive definite.) Nonconvex QPs, in which G is an indefinite matrix, can be more challenging because they can have several stationary points and local minimal. In this chapter we focus primarily on convex quadratic programs.

1.1 Applications

Portfolio optimization. This is formulated as solving the following optimization problem:

$$\max \ \boldsymbol{x}^T\mu - \kappa\boldsymbol{x}^T G\boldsymbol{x}, \qquad \text{subject to} \quad \sum_{i=1}^{n} x_i = 1, \boldsymbol{x} \geq 0.$$

Intuitively, we would like to find a portfolio for which the expected return $\boldsymbol{x}^T\mu$ is large while the variance $\boldsymbol{x}^T G\boldsymbol{x}$ is small.

MAP inference and Shape matching.

$$\max\ \boldsymbol{b}^T\boldsymbol{x} + \frac{1}{2}\boldsymbol{x}^T Q\boldsymbol{x}, \quad \sum_{j=1}^{m} x_{ij} = 1,\ 1 \le i \le n, \qquad \boldsymbol{x} \ge 0.$$

Here vector $\boldsymbol{b}$ encodes the first-order potential and Q encodes the second-order potential.

1.2 Equality-Constrained Quadratic Programs

For simplicity, we write the equality constraints in matrix form and state the equality- constrained QP as follows:

$$\min_{\boldsymbol{x}} \quad q(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^T G\boldsymbol{x} + \boldsymbol{x}^T\boldsymbol{c} \tag{4}$$

$$\text{subject to} \quad A\boldsymbol{x} = \boldsymbol{b}. \tag{5}$$

The first-order necessary conditions for $\boldsymbol{x}^\star$ to be a solution of (5) state that there is a vector $\lambda^\star$ such that the following system of equations is satisfied:

$$\begin{bmatrix} G & -A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{x}^\star \\ \lambda^\star \end{bmatrix} = \begin{bmatrix} -\boldsymbol{c} \\ \boldsymbol{b} \end{bmatrix}. \tag{6}$$

These conditions are a consequence of the general result for first-order optimality conditions. We call $\lambda^\star$ the vector of Lagrange multipliers. The system (6) can be rewritten in a form that is useful for computation by expressing $\boldsymbol{x}^\star$ as $\boldsymbol{x}^\star - \boldsymbol{x} + \boldsymbol{p}$, where $\boldsymbol{x}$ is some estimate of the solution and $\boldsymbol{p}$ is the desired step. By introducing this notation and rearranging the equations, we obtain

$$\begin{bmatrix} G & -A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} -\boldsymbol{p} \\ \lambda^\star \end{bmatrix} = \begin{bmatrix} \boldsymbol{g} \\ \boldsymbol{h} \end{bmatrix}. \tag{7}$$

where

$$\boldsymbol{h} = A\boldsymbol{x} - \boldsymbol{b}, \quad \boldsymbol{g} = \boldsymbol{c} + G\boldsymbol{x}, \quad \boldsymbol{p} = \boldsymbol{x}^\star - \boldsymbol{x}. \tag{8}$$

The matrix in (7) is called the Karush-Kuhn-Tucker (KKT) matrix, and the following result gives conditions under which it is nonsingular. We denote the $n \times (n-m)$ matrix whose columns are a basis for the null space of A. That is, Z has full rank and satisfies $A \cdot Z = 0$.

Lemma 1.1. *Let A have full row rank, and assume that the reduced-Hessian matrix $Z^T GZ$ is positive definite. Then the KKT matrix*

$$\begin{bmatrix} G & -A^T \\ A & 0 \end{bmatrix} \tag{9}$$

is nonsingular, and hence there is a unique vector pair $(\boldsymbol{x}^\star, \lambda^\star)$ satisfying (6).

Theorem 1.1. *Let A have full row rank and assume that the reduced-Hessian matrix $Z^T GZ$ is positive definite. Then the vector $\boldsymbol{x}^\star$ satisfying (6) is the unique global solution of (5).*

The KKT system can be solved using factorization method, e.g., LU factorization. Another option is to solve it using conjugate gradient methods.

1.3 Inequality-Constrained Problems

In the remainder of the chapter we discuss several classes of algorithms for solving convex quadratic programs that contain both inequality and equality constraints. Active-set methods have been widely used since the 1970s and are effective for small- and medium-sized problems. They allow for efficient detection of unboundedness and infeasibility and typically return an accurate estimate of the optimal active set. Interior-point methods are more recent, having become popular in the 1990s. They are well suited for large problems but may not be the most effective when a series of related QPs must be solved. We also study a special type of active-set methods called a gradient projection method, which is most effective when the only constraints in the problem are bounds on the variables.

Optimality Conditions for Inequality-Constrained Problems. As what we have discussed in the optimality conditions of constrained optimization techniques, the active set $\mathcal{A}(\boldsymbol{x}^\star)$ consists of the indices of the constraints for which equality holds at $\boldsymbol{x}^\star$:

$$\mathcal{A}(\boldsymbol{x}^\star) = \{i \in \mathcal{E} \cup \mathcal{I} | \boldsymbol{a}_i^T \boldsymbol{x}^\star = b_i\}. \tag{10}$$

By specializing the KKT conditions to this problem, we find that any solution $\boldsymbol{x}^\star$ of (3) satisfies the following first-order conditions, for some Lagrange multipliers $\lambda_i^\star, i \in \mathcal{A}(\boldsymbol{x}^\star)$:

$$\begin{aligned} G\boldsymbol{x}^\star + \boldsymbol{c} - \sum_{i \in \mathcal{A}(\boldsymbol{x}^\star)} \lambda_i^\star \boldsymbol{a}_i = 0,& \\ \boldsymbol{a}_i^T \boldsymbol{x}^\star = b_i, &\quad \forall i \in \mathcal{A}(\boldsymbol{x}^\star), \\ \boldsymbol{a}_i^T \boldsymbol{x}^\star \geq b_i, &\quad \forall i \in \mathcal{I} \setminus \mathcal{A}(\boldsymbol{x}^\star), \\ \lambda_i^\star \geq 0, &\quad i \in \mathcal{I} \cap \mathcal{A}(\boldsymbol{x}^\star). \end{aligned} \tag{11}$$

Theorem 1.2. *If $\boldsymbol{x}^\star$ satisfies the conditions (11) for some $\lambda_i^\star, i \in \mathcal{A}(\boldsymbol{x}^\star)$, and G is positive semidefinite, then $\boldsymbol{x}^\star$ is a global solution of (3).*

Proof. First of all, we have for any other feasible solution $\boldsymbol{x}$,

$$(\boldsymbol{x} - \boldsymbol{x}^\star)^T (G\boldsymbol{x}^\star + \boldsymbol{c}) = \sum_{i \in \mathcal{E}} \lambda_i^\star \boldsymbol{a}_i^T (\boldsymbol{x} - \boldsymbol{x}^\star) + \sum_{i \in \mathcal{A}(\boldsymbol{x}^\star) \cap \mathcal{I}} \lambda_i^\star \boldsymbol{a}_i^T (\boldsymbol{x} - \boldsymbol{x}^\star) \geq 0.$$

By elementary manipulation, we find that

$$q(\boldsymbol{x}) = q(\boldsymbol{x}^\star) + (\boldsymbol{x} - \boldsymbol{x}^\star)^T (G\boldsymbol{x}^\star + \boldsymbol{c}) + \frac{1}{2} (\boldsymbol{x} - \boldsymbol{x}^\star)^T G (\boldsymbol{x} - \boldsymbol{x}^\star) \geq q(\boldsymbol{x}^\star).$$

$\square$

1.4 Interior-Point Methods

The interior-point approach can be applied to convex quadratic programs through a simple extension of the linear-programming algorithms. For simplicity, we restrict our attention to convex quadratic programs with inequality constraints, which we write as follows:

$$\begin{aligned} \min_{\boldsymbol{x}} \quad & q(\boldsymbol{x}) = \frac{1}{2} \boldsymbol{x}^T G \boldsymbol{x} + \boldsymbol{x}^T \boldsymbol{c} \\ \text{subject to} \quad & A\boldsymbol{x} \geq \boldsymbol{b}, \end{aligned} \tag{12}$$

where G is symmetric and positive semidefinite and where the $m \times n$ matrix A and right-hand side $\boldsymbol{b}$ are defined by

$$A = [\boldsymbol{a}_i]_{i \in \mathcal{I}}, \qquad \boldsymbol{b} = [b_i]_{i \in \mathcal{I}}, \qquad \mathcal{I} = \{1, \cdots, m\}.$$

Rewriting the KKT conditions (7) in this notation, we obtain

$$G\boldsymbol{x} - A^T \lambda + \boldsymbol{c} = 0, \tag{13}$$

$$A\boldsymbol{x} - \boldsymbol{b} \geq 0, \tag{14}$$

$$(A\boldsymbol{x} - \boldsymbol{b})_i \lambda_i = 0, \quad i = 1, 2, \cdots, m, \quad \lambda \geq 0. \tag{15}$$

By introducing the slack vector $\boldsymbol{y} \geq 0$, we can rewrite these conditions as

$$G\boldsymbol{x} - A^T\lambda + \boldsymbol{c} = 0, \tag{16}$$

$$A\boldsymbol{x} - \boldsymbol{y} - \boldsymbol{b} = 0, \tag{17}$$

$$y_i\lambda_i = 0, \quad i = 1, 2, \cdots, m, \quad (\boldsymbol{y}, \lambda) \geq 0. \tag{18}$$