

1 Quadratic Programming

Now, we extend the optimization problem to a quadratic objective function and linear constraints (known as a quadratic program - QP). This type of program actually arises in many subproblems for general constrained optimization, such as sequential quadratic programming (SQP), augmented Lagrangian methods and interior point methods.

The general QP formulation is the following:

$$\min_x q(x) = \frac{1}{2}x^T Gx + x^T c \quad (1)$$

$$\text{subject to } a_i^T x = b_i, \quad i \in \mathcal{E} \quad (2)$$

$$a_i^T x \geq b_i, \quad i \in \mathcal{I} \quad (3)$$

We concern ourselves now with mainly convex QPs, where the Hessian matrix of G is positive semidefinite.

1.1 Equality-constrained QPs

First, we study the case of having only equality constraints in the QP. That is, the form of the optimization problem is:

$$\min_{x \in \mathbb{R}^n} q(x) = \frac{1}{2}x^T Gx + x^T c \quad (4)$$

$$\text{subject to } Ax = b \quad (5)$$

where A is a $m \times n$ Jacobian of constraints matrix with $m \leq n$ (wide matrix), whose rows are a_i^T . Here $i \in \mathcal{E}$ and $\mathcal{I} = \emptyset$. $b \in \mathbb{R}^m$.

We assume that A has full row-rank (i.e. $\text{rank}(A) = m$), so that the constraints are consistent. First, we can analyze the first-order necessary conditions for x^* through the KKT conditions, such that there is a vector λ^* that satisfies the following system of equations:

$$\begin{pmatrix} G & -A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} x^* \\ \lambda^* \end{pmatrix} = \begin{pmatrix} -c \\ b \end{pmatrix} \quad (6)$$

Usually, this is rewritten with $x^* = x + p$, where x is an estimate of the solution and p is the desired step. This change of variables is important because it allows us to plug in any initial estimate x, and then compute the desired offset to the optimal solution. This might be more efficient than say, directly inverting the KKT matrix.

$$K = \begin{pmatrix} G & A^T \\ A & 0 \end{pmatrix} \begin{pmatrix} -p \\ \lambda^* \end{pmatrix} = \begin{pmatrix} g \\ h \end{pmatrix} \quad (7)$$

which is also known as the **KKT matrix**. Here:

$$h = Ax - b$$

and $g = Gx + c$ and $p = x^* - x$

Note that there is a unique solution for (x^*, λ^*) , when the KKT matrix is nonsingular. The next lemma gives conditions on the original constraint matrix, A , and the reduced-Hessian matrix such that the KKT matrix is non-singular. We define a matrix, Z , as a $n \times (n - m)$ matrix whose columns comprise of the null space of A .

[Sufficient conditions that KKT matrix is nonsingular] If A has full row rank, and the reduced-Hessian matrix $Z^T G Z$ is positive definite, then the KKT matrix is nonsingular.

In addition to having a unique solution, the vector x^* is a unique global solution under the assumptions of the above lemma.

[Sufficient conditions that solution to KKT matrix problem is unique global solution] If A has full row rank, and the reduced-Hessian matrix $Z^T G Z$ is positive definite, then the vector x^* satisfying the system of equations in 7 is a global unique solution.

Given that we have a nice theoretical properties where solving the KKT system for a nice-enough A and G matrix, then we are guaranteed a unique and global solution to our optimization problem. Now, we are interested in **efficient** methods for solving the KKT system.

1.1.1 Inertia of the KKT matrix

First, let us define the inertia of a symmetric matrix to be:

[inertia] $\text{inertia}(K) = (n_+, n_-, n_0)$ where n_+, n_-, n_0 are the number of positive, negative and zero eigenvalues respectively.

Note that when $m \geq 1$, the KKT matrix is always indefinite. We can characterize the inertia of the KKT matrix:

[Inertia of the KKT matrix] Let K be the KKT matrix and suppose A has rank m . Then the inertia of K is:

$$\text{inertia}(K) = \text{inertia}(Z^T G Z) + (m, m, 0)$$

So if $Z^T G Z$ is positive definite, then $\text{inertia}(K) = (n, m, 0)$.

Thus this tells us that the KKT system is always indefinite when at least one constraint. To solve this system of equations, there are now a few approaches one can take.

1.1.2 Matrix Factorization Methods for Solving the KKT System

Direct methods involve matrix factorization either by: i) symmetric indefinite factorization ii) the Schur-complement method iii) Null-space method

1.1.3 Iterative Methods for Solving the KKT System

When you have a very large system, or you have a desire to implement parallelization, then an iterative approach can be taken.

1.2 Inequality Constrained QPs

When we introduce inequality constraints, now the problem becomes more complex. One can use active-set methods, or interior point methods. Before we explore methods for solving these class of problems, we review the optimality conditions for inequality-constrained QPs. First, we write the Lagrangian for this problem and then write the KKT conditions for this problem:

KKT conditions of inequality-constrained QP

$$Gx^* + c - \sum_{i \in A(x^*)} \lambda_i^* a_i = 0 \quad (8)$$

$$a_i^T x^* = b_i \quad \text{for all } i \in A(x^*) \quad (9)$$

$$a_i^T x^* \geq b_i \quad \text{for all } i \in \mathcal{I} / A(x^*) \quad (10)$$

$$\lambda_i^* \geq 0 \quad \text{for all } i \in \mathcal{I} \cap A(x^*) \quad (11)$$

Similar to the equality constrained problem, if we assume that G is positive semidefinite, and a feasible point satisfies the conditions above, then we can show this proposed point is a global solution.

[Sufficient conditions for a global solution] If x^* satisfies the 8 conditions for some λ_i^* for $i \in A(x^*)$ (that is constrain i is in the active set), and G is PSD, then x^* is a global solution of the QP. [Uniqueness] If in addition, G is PD, then x^* is a global unique solution of the QP.

1.3 Active set methods for convex QPs

We have seen that equality-constrained QPs are easily analyzed and solved with matrix factorization methods that essentially attempt to invert the KKT matrix to provide a solution.

We consider active-set methods here for the case of a convex QP (i.e. when the G matrix is PSD). If we knew the active set at the optimal x^* ; that is $A(x^*)$, then we can find solutions using matrix factorization in the equality-constrained problem.

However, since we do not **explicitly** know which constraints are active at the optimal solution, then we must also in tandem, determine the components of this set in a solver. As a reminder, the active set methods derive from the KKT conditions that stipulate that an inequality constraint either is active, or its Lagrange multiplier is zero (and thus does not affect the objective function).

We will explore primal active-set methods, where we solve an equality-constrained QP as a subproblem based on the current **active working set**. The equality constraints are the actual equality constraints of the problem and the

inequality constraints of the problem that are in the working set. Thus at each iterate, we solve the following QP subproblem:

$$\min_p \frac{1}{2} p^T G p + g_k^T p \quad (12)$$

$$\text{subject to } a_i^T p = 0, \quad i \in \mathcal{W}_k \quad (13)$$

where $g_k = Gx_k + c$ and $p = x - x_k$. This subproblem will produce a solution, p_k . The p_k variable tells us at iterate k , how much to move along this direction from our x_k vector. We move in that direction a fraction of p_k that depends on where all constraints are met (including ones not in the working set). That is:

$$x_{k+1} = x_k + \alpha_k p_k$$

where $\alpha_k \in [0, 1]$ is our step size at iteration k . How might we choose α_k in a theoretically supported manner?

1.3.1 Optimal step size in QP subproblem

We can derive an explicit formula for α_k by considering what happens to the constraints that are not in the working set. If $a_i^T p_k \geq 0$ for some $i \notin W_k$, then for all $\alpha_k \geq 0$, we have the following inequality:

$$a_i^T (x_k + \alpha_k p_k) \geq a_i^T x_k \geq b_i$$

The constraint i will be satisfied for non-negative α_k . Therefore, α_k is upper-bounded as:

$$\alpha_k \leq \frac{b_i - a_i^T x_k}{a_i^T p_k}$$

Since we want to decrease the objective, q , as much as possible, we want α_k as large as possible in its possible interval $[0, 1]$, while remaining feasible.

$$\alpha_k = \min(1, \min_{i \notin W_k, a_i^T p_k < 0} \frac{b_i - a_i^T x_k}{a_i^T p_k})$$

Thus, if $\alpha_k \neq 1$, then the constraints, at which the minimum is achieved, then those $i \notin W_k$ are called the blocking constraints. That is, these constraints block the program from taking a further step in the direction, p_k . We can continue these iterations, and construct a new working set, where W_{k+1} is W_k plus one of the blocking constraints from W_k . Then one can arrive at a solution, when all blocking constraints are added, and the subproblem as a solution $p = 0$.

1.3.2 Termination and Summary of the Algorithm

However, with inequality constraints, by the KKT conditions, we know that the corresponding Lagrange multipliers must be non-negative. If we find a corresponding $i \in \hat{W} \cap \mathcal{I}$, such that $\hat{\lambda}_i < 0$, then we can actually **decrease the objective** further by just removing that constraint.

The following summarizes how the active-set QP proceeds.

1. Initialize a working set, W_0
2. Compute the first-order direction to decrease the objective, p_k
3. Compute the corresponding step size to take in this direction, α_k
4. Add any blocking constraints in to form the new working set, W_k
5. Repeat the program
6. When converged, check the signage of the Lagrange multipliers of the inequality constraints
7. Remove constraints if necessary and repeat program

When we arrive at a solution, characterized by $p_k = 0$, then we have a KKT point. Furthermore, by the earlier theorem, if G is PD, then it is a unique global solution. Thus the value of the direction vector tell us our termination condition (i.e. 0 or not 0).

1.3.3 Initialization of the QP - Initial Estimate and Initial Working Set

Now, that we have examined the theoretical properties of the active-set method for convex QPs, we are interested in examining the optimal way to initialize this program. There are two things to determine: i) initial feasible point and ii) initial working set.

An initial feasible point can be given by the Phase I approach in linear programming.

To obtain the initial working set, W_0 , one can simply take a linearly independent subset of the active constraints at the solution of the feasibility problem. Note that when selecting the initial working set, adding a blocking constraint and deleting a constraint cannot introduce linear dependence, so the initialization with a linearly independent constraint set ensures linear independence.

In general, this is a difficult problem.

2 Sequential Quadratic Program

These can be seen as active set methods for non-linear programming. The SQP is an effective method for nonlinearly constrained optimization, which generates

steps by solving QP subproblems. We mainly consider active-set methods, which are now capable of handling significant nonlinearities in the constraints. Again, the objective program that we consider is:

$$\min_{x \in \mathbb{R}^n} f(x) \quad s.t. \quad c_i(x) = 0 \quad i \in \mathcal{E}, \quad c_i(x) \geq 0 \quad i \in \mathcal{I}$$

where now, f and c are assumed to be smooth functions (i.e. continuously differentiable N times). However, we do not **assume** that c is a linear function in x . There are two approaches to such classes of problems:

1. Inequality Constrained Quadratic Program (IQP): Here, we compute an estimate of the active set and the step direction at the same time.
2. EQP: here first estimate the active set, and then solve a quadratic program with equality constraints restricted to the active set.

To understand SQP methods, we will need to:

1. develop local methods for computing the step
2. line search and trust-region methods to achieve convergence from remote starting points.

First, let us explore the local SQP with equality constraints only.

2.1 Local SQP For Equality Constraints

We consider only equality constraints first. The general approach is to form a KKT matrix and then use Newton's method for solving a system of equations. We know that the Lagrangian function for this problem is:

$$L(x, \lambda) = f(x) - \lambda^T c(x)$$

We say $A(x)^T = (\nabla c_1(x), \dots, \nabla c_m(x))$ is the Jacobian matrix of the constraints that is a $n \times m$ matrix (n dimensions and m constraints). The first-order necessary KKT conditions for optimality can be written as a system of $n+m$ equations and $n+m$ unknowns.:

$$F(x, \lambda) = \begin{pmatrix} \nabla f(x) - A(x)^T \lambda \\ c(x) \end{pmatrix} = 0$$

This stems from the KKT conditions. In order to solve this with nonlinearities, we can use Newton's method by:

1. Computing the Jacobian of $F(x, \lambda)$
2. Compute the Newton step for (x_k, λ_k)
3. Plug in (p_k, p_λ) to the Newton-KKT system

The Jacobian of F is:

$$F'(x, \lambda) = \begin{pmatrix} \nabla_{xx}^2 L(x, \lambda) & -A(x)^T \\ A(x) & 0 \end{pmatrix}$$

and the Newton step is given by:

$$x_{k+1} = x_k + p_k, \quad \lambda_{k+1} = \lambda_k + p_\lambda$$

where (p_k, p_λ) are obtained by solving the Newton-KKT system:

$$\begin{pmatrix} \nabla_{xx}^2 L(x, \lambda) & -A_k^T \\ A_k & 0 \end{pmatrix} \begin{pmatrix} p_k \\ p_\lambda \end{pmatrix} = \begin{pmatrix} -\nabla f_k + A_k^T \lambda_k \\ c_k \end{pmatrix}$$

As seen before in the linear QP, and by properties of block matrices, we know that the Newton step is well-defined with a unique solution for the step if the KKT matrix is nonsingular. This KKT matrix is nonsingular if:

1. $A(x)$ the Jacobian matrix has full row rank (no redundant constraints)
2. The Hessian of the Lagrangian is positive definite on the tangent space of the constraints

2.2 QP SubProblem

Alternatively to the Newton method update, we can model the problem as a "sequential" QP problem, where the iterates are computed as a solution of a QP. The SQP fraamework is more practical and can be extended to inequality-constraints.

By our earlier analysis, we saw that in general solving inequality-constraint QPs involve some "guess" of the active set of constraints, and then an update. In this nonlinear inequality QP, we will see that if we can guess correctly the active set of constraints for the nonlinear problem, then we will get rapid Newton-like convergence to a local solution.

[Robinson theorem for rapid convergence of SQP for nonlinear problems] Suppose x^* is a local solution for our problem at which the KKT conditions are satisfied for some λ^* . We assume the LICQ conditions. In addition, we assume strict complementarity and second-order sufficient conditions hold at (x^*, λ^*) .

Then if (x_k, λ_k) are sufficiently close to (x^*, λ^*) , there is a local solution of the QP subproblem whose active set A_k is the same as the active set $A(x^*)$ of the nonlinear program.

This is a qualitative theoretical result that informs us that given a **good iterate that is "close" to our optimal solution**, then our working active set will converge to the true active set of the problem at a **local solution**.

2.3 Trust-Region SQP

Trust-region methods to solve SQP are good when one does not have the Hessian matrix, $\nabla_{xx}^2 L_k$, as positive definite. That is, **there is some level of control over the singularities of the Jacobian and Hessian matrices** and there is a mechanism for enforcing global convergence. A way to formulate the trust-region SQP is to add a trust-region constraint subproblem:

$$\begin{aligned} \min_p \quad & f_k + \nabla f_k^T p + \frac{1}{2} p^T \nabla_{xx}^2 L_k p \\ \text{subject to} \quad & \nabla c_i(x_k)^T p + c_i(x_k) = 0, \quad i \in \mathcal{E} \\ & \nabla c_i(x_k)^T p + c_i(x_k) \geq 0, \quad i \in \mathcal{I} \\ & \|p\| \leq \Delta_k \quad \text{Trust region of step} \end{aligned} \tag{14}$$

2.4 References

Nocedal, and Wright. Springer series in operations research and financial engineering Springer, New York, NY, 2. ed. edition, (2006).