5.2 Quadratic Programming

A quadratic program (QP) takes the form

$$\begin{array}{rl} \displaystyle\min_{\mathbf{x}\in\mathbb{R}^n} & f(\mathbf{x}) := \frac{1}{2}\mathbf{x}^T G\mathbf{x} + \mathbf{x}^T\mathbf{h} \\ \text{subject to} & A^T\mathbf{x} = \mathbf{b} \\ & C^T\mathbf{x} \geq \mathbf{d}, \end{array} \tag{5.20}$$

where $G \in \mathbb{R}^{n\times n}$ is symmetric and $A \in \mathbb{R}^{n\times p}$, $B \in \mathbb{R}^{n\times m}$.

We focus on this problem partly to make our life simpler, and partly because it plays an important role in the SQP method to be discussed in Section 5.3.

If G is positive semidefinite then (5.20) is convex and Theorem 5.12 applies.

5.2.1 Equality-Constrained QPs

In the absence of inequality constraints, (5.20) becomes

$$\begin{array}{rl} \displaystyle\min_{\mathbf{x}\in\mathbb{R}^n} & f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T G\mathbf{x} + \mathbf{x}^T\mathbf{h} \\ \text{subject to} & A^T\mathbf{x} = \mathbf{b}. \end{array} \tag{5.21}$$

Only the KKT conditions (1)+(2) are relevant, and they can be compactly expressed in the form of the linear system

$$\begin{pmatrix} G & -A \\ A^T & 0 \end{pmatrix}\begin{pmatrix} \mathbf{x}^* \\ \mu^* \end{pmatrix} = \begin{pmatrix} -\mathbf{h} \\ \mathbf{b} \end{pmatrix}. \tag{5.22}$$

One disadvantage of this formulation is that the matrix in (5.22) is not symmetric. This can be fixed in several ways, for example

$$\begin{pmatrix} G & A \\ A^T & 0 \end{pmatrix}\begin{pmatrix} -\mathbf{x}^* \\ \mu^* \end{pmatrix} = \begin{pmatrix} \mathbf{h} \\ -\mathbf{b} \end{pmatrix}. \tag{5.23}$$

The linear system (5.23) (and consequently also (5.22)) has a unique solution if and only if the involved matrix is invertible.

Lemma 5.13 *Let A have full column rank. Let the columns of Z be a basis for the null space of A^T and assume that Z^TGZ is positive definite. Then the so called* **KKT matrix**

$$K = \begin{pmatrix} G & A \\ A^T & 0 \end{pmatrix}$$

is invertible.

Proof. Suppose that $K\binom{\mathbf{x}}{\mathbf{y}} = 0$. Then $G\mathbf{x} = -A\mathbf{y}$ and $A^T\mathbf{x} = 0$. The latter relation states that $\mathbf{x}$ is in the null space of A^T and hence there is a vector $\mathbf{z}$ such that $\mathbf{x} = Z\mathbf{z}$. Inserting this into the first relation gives $\mathbf{z}^TZ^T\mathbf{G}Z\mathbf{z} = -\mathbf{z}^TZ^TA\mathbf{y} = 0$. From the positive definitness of $Z^T\mathbf{G}Z$ it follows that $\mathbf{z} = 0$ and therefore also $\mathbf{x} = 0$. Finally, the full column rank of A combined with $A\mathbf{y} = -G\mathbf{x} = 0$ yields $\mathbf{y}$. This shows that the null space of K is $\{0\}$. In other words, K is invertible. $\square$

The positive definiteness of the reduced Hessian Z^TGZ makes (5.21) essentially convex. Note that it is not required that the Hessian G itself is positive definite. A variant of Theorem 5.12 applies: One can show that the vector $\mathbf{x}^*$ satisfying (5.22) is the unique global solution of (5.21), provided that the conditions of Lemma 5.13 hold. See Theorem 16.2 in [NW].

It is important to note that the matrix K is always indefinite, unless A vanishes. We can therefore not apply solvers for symmetric positive definite linear systems (e.g., Cholesky factorization, CG method) directly for K, even when G is symmetric positive definite. However, similar to the technique used in the proof of Lemma 5.13 we can reduce this linear system to the matrix $Z^T\mathbf{G}Z$, to which, e.g., the CG method can be applied.

5.2.2 Active set methods

We now consider the general QP (5.20), for which the KKT conditions take the form

(1) $G\mathbf{x}^* - A\mu^* - C\lambda^* = -\mathbf{h}$;

(2) $A^T\mathbf{x}^* = \mathbf{b}$;

(3a) $C^T\mathbf{x}^* \geq \mathbf{d}$;

(3b) $\lambda^* \geq 0$;

(3c) $\lambda^* \odot (C^T\mathbf{x}^* - \mathbf{d}) = 0$;

where $\odot$ denotes the elementwise product. As already explained in Section 5.1.3, there is no need to require the LICQ, due to the linearity of the constraints.

In the following, we will only consider the case that G **is positive semidefinite**. This implies that (5.20) is convex and Theorem 5.12 applies. *If* the active set $\mathcal{A}(\mathbf{x}^*)$ was known, one could solve the KKT conditions above with the method discussed in Section 5.2.1 by including the active inequality constraints into the equality constraints and ignoring the inactive inequality constraints. *Active set methods* leverage this observation by maintaining a working set $\mathcal{W}_k \subset \{1, \dots, m\}$ that satisfies $\mathcal{W}_k \subseteq \mathcal{A}(\mathbf{x}_k)$ for the kth iterate $\mathbf{x}_k$.

Given a feasible iterate $\mathbf{x}_k$ and $\mathcal{W}_k$, active set methods proceed by considering the subproblem

$$\begin{array}{ll} \min\limits_{\mathbf{x}\in\mathbb{R}^n} & f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^TG\mathbf{x} + \mathbf{x}^T\mathbf{h} \\ \text{subject to} & A^T\mathbf{x} = \mathbf{b} \\ & C_{\mathcal{W}_k}^T\mathbf{x} = \mathbf{d}_{\mathcal{W}_k}, \end{array} \tag{5.24}$$

where $C_{\mathcal{W}_k}$ contains the rows of C corresponding to $\mathcal{W}_k$ and $\mathbf{d}_{\mathcal{W}_k} \in \mathbb{R}^{\#\mathcal{W}_k}$ contains the corresponding entries of $\mathbf{d}$. We reformulate the minimization problem (5.24) by defining the step

$$\mathbf{p}_k = \mathbf{x} - \mathbf{x}_k,$$

which is given as the solution of

$$\begin{array}{rl} \min\limits_{\mathbf{p}\in\mathbb{R}^n} & f(\mathbf{x}) = \frac{1}{2}\mathbf{p}^T G\mathbf{p} + \mathbf{p}^T(\mathbf{h} + G\mathbf{x}_k) \\ \text{subject to} & A^T\mathbf{p} = \mathbf{b} - A^T\mathbf{x}_k \\ & C_{\mathcal{W}_k}^T\mathbf{p} = \mathbf{d}_{\mathcal{W}_k} - C_{\mathcal{W}_k}^T\mathbf{x}_k, \end{array} \tag{5.25}$$

where we dropped constant terms in the object function. According to Section 5.2.1, see (5.23), the subproblem (5.25) can be addressed by solving the linear system

$$\begin{pmatrix} G & A & C_{\mathcal{W}_k} \\ A^T & 0 & 0 \\ C_{\mathcal{W}_k}^T & 0 & 0 \end{pmatrix} \begin{pmatrix} -\mathbf{p}_k \\ \mu_{k+1} \\ \lambda_{k+1} \end{pmatrix} = \begin{pmatrix} \mathbf{h} + G\mathbf{x}_k \\ A^T\mathbf{x}_k - \mathbf{b} \\ C_{\mathcal{W}_k}^T\mathbf{x}_k - \mathbf{d}_{\mathcal{W}_k} \end{pmatrix}. \tag{5.26}$$

By Lemma 5.13, we need to make sure that $\begin{pmatrix} A & C_{\mathcal{W}_k} \end{pmatrix}$ has full column rank. Strategies that guarantee this property are discussed in Section 16.5 of [NW].

Given the solution $\mathbf{p}_k$ of (5.26), the updated vector $\mathbf{x}_k + \mathbf{p}_k$ solves (5.24). More importantly,

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k\mathbf{p}_k$$

satisfies the constraints of (5.24) for any $\alpha_k \in \mathbb{R}$. Because $\mathbf{x}_k$ is feasible, it is always possible to choose α_k to be the largest value in $[0, 1]$ such that $\mathbf{x}_{k+1}$ is also feasible. Considering the ith row of the relation $C^T\mathbf{x} \geq \mathbf{d}$, it is clear that we only need to worry about rows for which $i \notin \mathcal{W}_k$ and $\mathbf{c}_i^T\mathbf{p}_k < 0$. It follows that

$$\alpha_k := \min\left\{1,\ \min_{\substack{i\notin\mathcal{W}_k \\ \mathbf{c}_i^T\mathbf{p}_k<0}} \frac{\mathbf{d}_i - \mathbf{c}_i^T\mathbf{x}_k}{\mathbf{c}_i^T\mathbf{p}_k}\right\}. \tag{5.27}$$

If $\alpha_k < 1$ then the step $\mathbf{p}_k$ is blocked by at least one constraint j not contained in $\mathcal{W}_k$. We continue the method by setting

$$\mathcal{W}_{k+1} := \mathcal{W}_k \cup \{j\}.$$

The above procedure stagnates when it encounters $\mathbf{p}_k = 0$, implying that $\mathbf{x}_k$ is already optimal with respect to the current working set $\mathcal{W}_k$. Let λ_{k+1}, μ_{k+1} denote the corresponding Lagrange multipliers, where we set the entries of λ_{k+1} to zero at indices not contained in $\mathcal{W}_k$. If $\lambda_{k+1} \geq 0$ then all KKT conditions for (5.20) are satisfied. Hence, we have found a global solution $\mathbf{x}_k$ and stop the method. If, on the other hand, there is an index $j \in \mathcal{W}_k$ with $\lambda_{k+1,j} < 0$ then this index is removed,

$$\mathcal{W}_{k+1} := \mathcal{W}_k \setminus \{j\},$$

and the method is continued. In summary, we obtain the algorithm below. Note that feasible starting points can be obtained by using techniques from linear programming, see [E,NW].

Algorithm 5.1. Active set method for convex QP.

Input: QP (5.20) with symmetric positive semidefinite G and feasible starting point $\mathbf{x}_0$.

Output: Global solution $\mathbf{x}^*$.

for $k = 0, 1, 2, \ldots$ **do**
Compute $\mathbf{p}_k$ such that $\mathbf{p}_k + \mathbf{x}_k$ solves (5.24), along with the corresponding Lagrange multipliers λ_{k+1}, μ_{k+1}.
if $\mathbf{p}_k = 0$ and $\lambda_{k+1} \geq 0$ **then**
Stop with $\mathbf{x}^* = \mathbf{x}_k$.
else if $\mathbf{p}_k = 0$ and $\lambda_{k+1} \ngeq 0$ **then**
Choose $j \leftarrow \arg\min_{j \in \mathcal{W}_k} \lambda_{k+1,j}$.
Set $\mathbf{x}_{k+1} \leftarrow \mathbf{x}_k, \mathcal{W}_{k+1} \leftarrow \mathcal{W}_k \setminus \{j\}$.
else
Compute α_k as in (5.27) and set $\mathbf{x}_{k+1} \leftarrow \mathbf{x}_k + \alpha_k \mathbf{p}_k$.
if $\alpha_k < 1$ **then**
Obtain $\mathcal{W}_{k+1}$ by adding blocking constraint to $\mathcal{W}_k$.
else
Set $\mathcal{W}_{k+1} \leftarrow \mathcal{W}_k$.
end if
end if
end for

One of the most remarkable properties of this algorithm is that it terminates after a *finite* number of steps, provided that $\alpha_k > 0$ for every nonzero search direction $\mathbf{p}_k$, see Pg. 157 in [NW]. In the rare case that $\alpha_k = 0$, the algorithm may run into a cycle. Tricks similar to the ones used in linear programming can be used to break the cycle.

5.2.3 Interior point methods

Together with the simplex method, interior point methods are one of the most frequently used solvers in linear programming. They have also turned out to be quite successful in addressing general nonconvex nonlinear optimization problems, see Chapter 19 in [NW]. In this lecture, we will restrict our coverage of this important class of methods to convex quadratic programs. Moreover, for simplicity, we will neglect equality constraints:

$$\begin{array}{rl} \displaystyle\min_{\mathbf{x} \in \mathbb{R}^n} & f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T G \mathbf{x} + \mathbf{x}^T \mathbf{h} \\ \text{subject to} & C^T \mathbf{x} \geq \mathbf{d}, \end{array} \tag{5.28}$$

where G is symmetric positive semidefinite. Let us recall that the KKT conditions take the form

(1) $G\mathbf{x} - C\lambda = -\mathbf{h}$;

(3a) $C^T \mathbf{x} \geq \mathbf{d}$;

(3b) $\lambda \geq 0$;

(3c) $\lambda \odot (C^T \mathbf{x} - \mathbf{d}) = 0$.

Thanks to convexity, these conditions are sufficient and necessary for a global solution of (5.28).

To avoid dealing with $C^T\mathbf{x} \geq \mathbf{d}$, we now introduce the so called **slack vector** $\mathbf{s} \geq 0$, yielding the equivalent conditions

(1) $G\mathbf{x} - C\lambda = -\mathbf{h}$;

(Sa) $C^T\mathbf{x} - \mathbf{s} = \mathbf{d}$;

(Sb) $\lambda \geq 0; \mathbf{s} \geq 0$;

(Sc) $\lambda \odot \mathbf{s} = 0$.

Interior point methods generate iterates $(\mathbf{x}, \lambda, \mathbf{s})$ that satisfy (Sb) *strictly*, that is, $\lambda > 0$ and $\mathbf{s} > 0$. Of course, this means that (Sc) cannot be satisfied exactly. To quantify how far we are off from complementarity, we consider the average of the entries in $\lambda \odot \mathbf{s}$ (which are all positive):

$$\mu := \frac{1}{m}\mathbf{s}^T\lambda. \tag{5.29}$$

The correspondingly *perturbed* KKT conditions can be written as a nonlinear system

$$F(\mathbf{x}, \mathbf{s}, \lambda, \sigma\mu) := \begin{pmatrix} G\mathbf{x} - C\lambda + \mathbf{h} \\ C^T\mathbf{x} - \mathbf{s} - \mathbf{d} \\ \mathcal{S}\Lambda\mathbf{e} - \sigma\mu\mathbf{e} \end{pmatrix} = 0, \tag{5.30}$$

where

$$\mathcal{S} = \mathrm{diag}(s_1, \ldots, s_m), \quad \Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_m), \quad \mathbf{e} = (1, \ldots, 1)^T.$$

The solutions of (5.30) define the so called **central path**. We aim to bring μ down to zero, while maintaining the positivity of $\mathbf{y}$ and $\mathbf{s}$ at the same time.

By considering μ fixed, Newton's method applied to (5.30) yields the linear system

$$\begin{pmatrix} G & 0 & -C \\ C^T & -I & 0 \\ 0 & \Lambda & \mathcal{S} \end{pmatrix} \begin{pmatrix} \triangle\mathbf{x} \\ \triangle\mathbf{s} \\ \triangle\lambda \end{pmatrix} = \begin{pmatrix} -\mathbf{r}_d \\ -\mathbf{r}_p \\ -\mathcal{S}\Lambda\mathbf{e} + \sigma\mu\mathbf{e} \end{pmatrix}, \tag{5.31}$$

with

$$\mathbf{r}_d = G\mathbf{x} - C\lambda + \mathbf{h}, \qquad \mathbf{r}_p = C^T\mathbf{x} - \mathbf{s} - \mathbf{d}.$$

Given a current iterate $(\mathbf{x}_k, \mathbf{s}_k, \lambda_k)$ with $\mathbf{s}_k > 0, \lambda_k > 0$, interior point methods proceed as follows to produce the next iterate:

1. Compute $\mu = \frac{1}{m}\mathbf{s}^T\lambda$ according to (5.29).
2. Choose the barrier parameter reduction factor $\sigma \in [0, 1]$.
3. Determine the Newton direction $(\triangle\mathbf{x}, \triangle\mathbf{s}, \triangle\lambda)$ from the current iteration by solving (5.31).

4. Choose a step length $\alpha > 0$ such that $\mathbf{x}_{k+1} := \mathbf{x}_k + \alpha\triangle\mathbf{x} > 0$ and $\mathbf{s}_{k+1} := \mathbf{s}_k + \alpha\triangle\mathbf{s} > 0$.

5. Set $\lambda_{k+1} = \lambda_k + \alpha\triangle\lambda$.

The convergence speed of this iteration crucially depends on the choice of σ in Step 2 and the choice of α in Step 4. We refer to the discussion in Section 16.6 of [NW].

5.3 Sequential Quadratic Programming (SQP)*

Sequential Quadratic Programming is one of the most successful techniques in dealing with general nonlinear constrained optimization problems. It generates steps by solving quadratic programming subproblems, using one of the algorithms discussed in the previous section.

5.3.1 Local SQP for equality constraints

To start simple, we first discuss the computation of steps for the equality-constrained problem

$$\begin{array}{rl} \min\limits_{\mathbf{x}\in\mathbb{R}^n} & f(\mathbf{x}) \\ \text{subject to} & h(\mathbf{x}) = 0. \end{array} \tag{5.32}$$

Let us recall that the Lagrangian for this problem is given by

$$\mathcal{L}(x,\lambda) = f(\mathbf{x}) - \mu^T h(\mathbf{x}),$$

leading to the KKT conditions

$$F(\mathbf{x},\mu) := \begin{pmatrix} \nabla f(\mathbf{x}) - h'(\mathbf{x})^T\mu \\ h(\mathbf{x}) \end{pmatrix} = 0.$$

The Newton method applied to this nonlinear system takes the form

$$\begin{pmatrix} \mathbf{x}_{k+1} \\ \mu_{k+1} \end{pmatrix} = \begin{pmatrix} \mathbf{x}_k \\ \mu_k \end{pmatrix} + \begin{pmatrix} \mathbf{p}_k \\ \mathbf{p}_{k,\mu} \end{pmatrix} \tag{5.33}$$

with the correction vector satisfying the linear system

$$\begin{pmatrix} \mathsf{H}_k & -h'(\mathbf{x}_k)^T \\ h'(\mathbf{x}_k) & 0 \end{pmatrix} \begin{pmatrix} \mathbf{p}_k \\ \mathbf{p}_{k,\mu} \end{pmatrix} = \begin{pmatrix} -\nabla f(\mathbf{x}_k) + h'(\mathbf{x}_k)^T\mu_k \\ -h(\mathbf{x}_k) \end{pmatrix} \tag{5.34}$$

Here, $\mathsf{H}_k := \mathcal{L}_{\mathbf{xx}}(\mathbf{x}_k,\mu_k)$ denotes the Hessian of $\mathcal{L}$ with respect to $\mathbf{x}$ at $(\mathbf{x}_k,\mu_k)$. By Lemma 5.13, the so called Newton-KKT system (5.34) is uniquely solvable at $(\mathbf{x},\mu)$ if

(A1) The constraint Jacobian $h'(\mathbf{x})$ has full row rank.

(A2) $\mathbf{d}^T\mathsf{H}_k\mathbf{d} > 0$ for all $\mathbf{d}$ such that $h'(\mathbf{x})\mathbf{d} = 0$.

Assumption (A1) is our good old LICQ, see Definition 5.8. For $(\mathbf{x}, \mu)$ sufficiently close to a local solution $(\mathbf{x}^*, \mu^*)$, Assumption (A2) follows from a second-order sufficiency condition for $(\mathbf{x}^*, \mu^*)$ that has been discussed in the exercises, see also [NW]. Since (A1) and (A2) guarantee the invertibility of the Jacobian, standard results for Newton methods imply local quadratic convergence to a local solution $(\mathbf{x}^*, \mu^*)$ for which LICQ and the second-order sufficiency condition hold.

We now present an alternative way to derive (5.34), which admits a more straightforward extension to inequality constraints. Around the current iterate $(\mathbf{x}_k, \mu_k)$, the nonlinear problem (5.32) is replaced by the quadratic program

$$\begin{array}{rl} \min\limits_{\mathbf{p}\in\mathbb{R}^n} & f(\mathbf{x}_k) + \nabla f(\mathbf{x}_k)^T\mathbf{p} + \frac{1}{2}\mathbf{p}^T\mathsf{H}_k\mathbf{p} \\ \text{subject to} & h'(\mathbf{x}_k)\mathbf{p} + h(\mathbf{x}_k) = 0. \end{array} \tag{5.35}$$

Note that the constant term $f(\mathbf{x}_k)$ has no effect on the minimizer. By the discussion in Section 5.2.1, see in particular (5.22)[7], the solution $\mathbf{p}_k$ of the QP (5.35) satisfies the linear system

$$\begin{pmatrix} \mathsf{H}_k & -h'(\mathbf{x}_k) \\ h'(\mathbf{x}_k)^T & 0 \end{pmatrix} \begin{pmatrix} \mathbf{p}_k \\ \mu_{k+1} \end{pmatrix} = \begin{pmatrix} -\nabla f(\mathbf{x}_k) \\ -h(\mathbf{x}_k) \end{pmatrix}. \tag{5.36}$$

This linear system is uniquely solvable if assumptions (A1) and (A2) are satisfied.

Using $\mu_{k+1} = \mu_k + \mathbf{p}_{k,\mu}$, it turns out that (5.36) is identical with (5.34). Hence both approaches, (i) applying Newton to the KKT conditions and (ii) locally solving a QP, produce the same iterates!

5.3.2 Local SQP for equality and inequality constraints

The local SQP method (5.35) is easily extended to the general constrained optimization problem (5.1):

$$\begin{array}{rl} \min\limits_{\mathbf{p}\in\mathbb{R}^n} & \nabla f(\mathbf{x}_k)^T\mathbf{p} + \frac{1}{2}\mathbf{p}^T\mathsf{H}_k\mathbf{p} \\ \text{subject to} & g'(\mathbf{x}_k)\mathbf{p} + g(\mathbf{x}_k) \geq 0 \\ & h'(\mathbf{x}_k)\mathbf{p} + h(\mathbf{x}_k) = 0. \end{array} \tag{5.37}$$

The solution $\mathbf{p}_k$ to this QP can be obtained with the methods discussed in Section 5.2, resulting in the next iterate $\mathbf{x}_{k+1} \leftarrow \mathbf{x}_k + \mathbf{p}_k$, along with the correspondingly updated Lagrange multipliers λ_{k+1}, μ_{k+1}. It can be shown that the sequence resulting $(\mathbf{x}_k, \lambda_k, \mu_k)$ converges locally quadratically to a triple $(\mathbf{x}^*, \lambda^*, \mu^*)$ satisfying the KKT conditions.

5.3.3 Globalized SQP

To perform line search (or any other kind of strategy ensuring global convergence), we not only have to take the value of the objective function into account but also

[7] Be careful, the choice of notation in this section and in Section 5.2.1 do not match.

the satisfaction of the constraints. This is the purpose of **merit functions**. One popular choice is the ℓ_1 penalty function defined by

$$\phi_\beta(\mathbf{x}) = f(\mathbf{x}) + \beta \sum_{i=1}^m \max\{0, -g_i(\mathbf{x})\} + \beta \sum_{i=1}^p |h_i(\mathbf{x})|$$

for some parameter $\beta > 0$. Defining the vector $(g(\mathbf{x}))_-$ componentwise by $(g_i)_- = \max\{0, -g_i(\mathbf{x})\}$ and using the usual ℓ_1 vector norm, we can express this penalty function in more compact form as

$$\phi_\beta(\mathbf{x}) = f(\mathbf{x}) + \beta\|(g(\mathbf{x}))_-\|_1 + \beta\|h(\mathbf{x})\|_1. \tag{5.38}$$

Due to the presence of the maximum and the absolute value in the definition, the function ϕ_β is not differentiable everywhere. Instead, we will work with the (one-sided) directional derivative *[dérivée directionnelle au sens de Dini]*, which is defined as

$$D_+(\phi_\beta(\mathbf{x}))[\mathbf{p}] = \lim_{\substack{\alpha\to 0\\ \alpha>0}} \frac{\phi_\beta(\mathbf{x}+\alpha\mathbf{p}) - \phi_\beta(\mathbf{x})}{\alpha}.$$

for a direction $\mathbf{p} \in \mathbb{R}^n$. This derivative always exists and significantly simplifies if we only consider directions produced by local SQP.

Lemma 5.14 *Let f, g, h be continuously differentiable and let $\mathbf{p}_k$ satisfies the constraints of the QP* (5.37). *Then*

$$D_+(\phi_\beta(\mathbf{x}_k))[\mathbf{p}_k] = \nabla f(\mathbf{x}_k)^T\mathbf{p}_k - \beta\|h(\mathbf{x}_k)\|_1 - \beta \sum_{g_i(\mathbf{x}_k)<0} \nabla g_i(\mathbf{x}_k)^T\mathbf{p}_k. \tag{5.39}$$

Proof. We treat each term in (5.38) separately. Since f is differentiable, we have

$$D_+(f(\mathbf{x}_k))[\mathbf{p}_k] = \nabla f(\mathbf{x}_k)^T\mathbf{p}_k. \tag{5.40}$$

For the second term, we first note that the differentiability of g implies

$$\|(g(\mathbf{x}_k+\alpha\mathbf{p}_k))_-\|_1 - \|g(\mathbf{x}_k)\|_1 = \|(g(\mathbf{x}_k)+\alpha g'(\mathbf{x}_k)\mathbf{p}_k))_-\|_1 - \|(g(\mathbf{x}_k))_-\|_1 + o(\alpha)$$

For $g_i(\mathbf{x}_k) > 0$, we have $(g(\mathbf{x}_k))_- = (g_i(\mathbf{x}_k) + \alpha\nabla g_i(\mathbf{x}_k)^T\mathbf{p}_k))_- = 0$ for sufficiently small α and hence $D_+((g_i(\mathbf{x}_k))_-)[\mathbf{p}_k] = 0$. For $g_i(\mathbf{x}_k) < 0$, we readily obtain $D_+((g_i(\mathbf{x}_k))_-)[\mathbf{p}_k] = -\nabla g_i(\mathbf{x}_k)^T\mathbf{p}_k$. For $g_i(\mathbf{x}_k) = 0$, we have

$$(g_i(\mathbf{x}_k) + \alpha\nabla g_i(\mathbf{x}_k)^T\mathbf{p}_k))_- - (g_i(\mathbf{x}_k))_- = \alpha(\nabla g_i(\mathbf{x}_k)^T\mathbf{p}_k))_- = 0,$$

where we used $\nabla g_i(\mathbf{x}_k)^T\mathbf{p}_k \geq -g(\mathbf{x}_k) > 0$ in the last step. This follows from the constraint $g'(\mathbf{x}_k)\mathbf{p}_k + g(\mathbf{x}_k) \geq 0$ in (5.37). In summary, we obtain

$$D_+(\|(g(\mathbf{x}_k))_-\|_1)[\mathbf{p}_k] = -\sum_{g_i(\mathbf{x}_k)<0} \nabla g_i(\mathbf{x}_k)^T\mathbf{p}_k. \tag{5.41}$$

For the third term, the differentiability of h implies

$$\begin{aligned}\|h(\mathbf{x}_k+\alpha\mathbf{p}_k)\|_1 - \|h(\mathbf{x}_k)\|_1 &= \|h(\mathbf{x}_k) + \alpha h'(\mathbf{x}_k)\mathbf{p}_k)\|_1 - \|h(\mathbf{x}_k)\|_1 + o(\alpha)\\ &= (1-\alpha)\|h(\mathbf{x}_k)\|_1 - \|h(\mathbf{x}_k)\|_1 + o(\alpha)\\ &= -\alpha\|h(\mathbf{x}_k)\|_1 + o(\alpha)\end{aligned}$$

for $\alpha \leq 1$, where we used the relation $h'(\mathbf{x}_k)\mathbf{p}_k = -h(\mathbf{x}_k)$ from the constraints in (5.37). Hence,

$$D_+(\|h(\mathbf{x}_k)\|_1)[\mathbf{p}_k] = \|h(\mathbf{x}_k)\|_1. \tag{5.42}$$

Combining (5.40), (5.41), and (5.42) concludes the proof. □

Provided that LICQ holds, a local solution of the QP will give rise to a triple $(\mathbf{p}_k, \lambda_{k+1}, \mu_{k+1})$ satisfying the KKT conditions

(1) $\mathsf{H}_k \mathbf{p}_k - g'(\mathbf{x}_k)^T \lambda_{k+1} - h'(\mathbf{x}_k)^T \mu_{k+1} = -\nabla f(\mathbf{x}_k)$;

(2) $h'(\mathbf{x}_k)\mathbf{p}_k + h(\mathbf{x}_k) = 0$;

(3a) $g'(\mathbf{x}_k)\mathbf{p}_k + g(\mathbf{x}_k) \geq 0$;

(3b) $\lambda_{k+1} \geq 0$;

(3c) $\lambda_{k+1} \odot (g'(\mathbf{x}_k)\mathbf{p}_k + g(\mathbf{x}_k)) = 0$.

As the following theorem shows, these KKT conditions together with the positive definiteness of the Hessian imply that p_k is a descent direction for the merit function ϕ_β, provided that β is sufficiently large.

Theorem 5.15 *Let f, g, h be continuously differentiable and let $(\mathbf{p}_k, \lambda_{k+1}, \mu_{k+1})$ satisfy the KKT conditions above. Then*

$$D_+(\phi_\beta(\mathbf{x}_k))[\mathbf{p}_k] \leq -\mathbf{p}_k^T \mathsf{H}_k \mathbf{p}_k,$$

provided that $\beta \geq \max\{\|\lambda_{k+1}\|_\infty, \|\mu_{k+1}\|_\infty\}$.

Proof. The KKT conditions (1) and (2) give

$$\begin{aligned}
\nabla f(\mathbf{x}_k)^T \mathbf{p}_k &= -\mathbf{p}_k^T \mathsf{H}_k \mathbf{p}_k + \lambda_{k+1}^T g'(\mathbf{x}_k)\mathbf{p}_k + \mu_{k+1}^T h'(\mathbf{x}_k)\mathbf{p}_k \\
&= -\mathbf{p}_k^T \mathsf{H}_k \mathbf{p}_k + \lambda_{k+1}^T g'(\mathbf{x}_k)\mathbf{p}_k - \mu_{k+1}^T h(\mathbf{x}_k).
\end{aligned}$$

The middle term is handled by the KKT conditions (3a)–(3c), which imply

$$\begin{aligned}
\lambda_{k+1}^T g'(\mathbf{x}_k)\mathbf{p}_k &= \sum_{g_i(\mathbf{x}_k)<0} \lambda_{k+1,i} \nabla g_i(\mathbf{x}_k)^T \mathbf{p}_k + \sum_{g_i(\mathbf{x}_k)\geq 0} \lambda_{k+1,i} \nabla g_i(\mathbf{x}_k)^T \mathbf{p}_k \\
&= \sum_{g_i(\mathbf{x}_k)<0} \lambda_{k+1,i} \underbrace{\nabla g_i(\mathbf{x}_k)^T \mathbf{p}_k}_{>0} - \sum_{g_i(\mathbf{x}_k)\geq 0} \lambda_{k+1,i} g_i(\mathbf{x}_k) \\
&\leq \beta \sum_{g_i(\mathbf{x}_k)<0} \nabla g_i(\mathbf{x}_k)^T \mathbf{p}_k.
\end{aligned}$$

Note that we also have $-\mu_{k+1}^T h(\mathbf{x}_k) \leq \beta \|h(\mathbf{x}_k)\|_1$.

Combining these relations with the result of Lemma 5.14 gives

$$\begin{aligned}D_+(\phi_\beta(\mathbf{x}_k))[\mathbf{p}_k] &= \nabla f(\mathbf{x}_k)^T\mathbf{p}_k - \beta\|h(\mathbf{x}_k)\|_1 - \beta \sum_{g_i(\mathbf{x}_k)<0} \nabla g_i(\mathbf{x}_k)^T\mathbf{p}_k \\ &\leq -\mathbf{p}_k^T \mathsf{H}_k \mathbf{p}_k - \mu_{k+1}^T h(\mathbf{x}_k) - \beta\|h(\mathbf{x}_k)\|_1 \\ &\leq -\mathbf{p}_k^T \mathsf{H}_k \mathbf{p}_k,\end{aligned}$$

which concludes the proof. □

None of the derivations above made use of the fact that H_k is the Hessian of the Lagrangian. Hence, the result of Theorem 5.15 remains valid for an arbitrary symmetric matrix H_k. Therefore the ideas for quasi-Newton methods from Section 4.2.5 carry over to constrained optimization problems. Again, it is important to maintain the positive definitness H_k to guarantee that $\mathbf{p}_k$ is a descent direction under the conditions of Theorem 5.15.

Finally, we provide the complete algorithms

Algorithm 5.2. SQP with Armijo line search.

Input: Functions f, g, h, starting vectors $\mathbf{x}_0, \lambda_0, \mu_0$ and parameters $\alpha > 0$, $\beta > 0$, $c_1 > 0$.

Output: Vector $\mathbf{x}_k$ approximating stationary point.

1: **for** $k = 0, 1, 2, \ldots$ **do**
2: Determine a solution $\mathbf{p}_k$ of the QP (5.35) along with Lagrange multipliers λ, μ.
3: Set $\mathbf{p}_\lambda = \lambda - \lambda_k$, $\mathbf{p}_\mu = \mu - \mu_k$.
4: Determine the largest number $\alpha_k \in \{1, 2^{-1}, 2^{-2}, \ldots\}$ such that

$$\phi_\beta(\mathbf{x}_k + \alpha_k\mathbf{p}_k) - \phi_\beta(\mathbf{x}_k) \leq c_1\alpha_k D_+(\phi_\beta(\mathbf{x}_k))[\mathbf{p}_k]$$

5: Set $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k\mathbf{p}_k$, $\lambda_{k+1} = \lambda_k + \alpha_k\mathbf{p}_\lambda$, $\mu_{k+1} = \mu_k + \alpha_k\mathbf{p}_\mu$.
6: **end for**

Algorithm 5.3.3 is the most basic variant of a globalized SQP algorithm. As already mentioned above, in practice one normally uses quasi-Newton techniques (such as BFGS) to avoid the computation of the Hessian. A number of additional subtleties have to be taken into account when implementing Algorithm 5.3.3. For example, it may happen that one QP (5.35) appearing in the course of Algorithm 5.3.3 does not admit a solution, for example because the feasible set is empty. One possible way out of such a situation is to relax the constraints of the QP (5.35). We refer to Chapter 18 in [NW] for this and other details on the SQP algorithm.