Chapter 11

Sequential quadratic programming methods

The material of this chapter is a quick, somewhat customized summary of the long Chapter 18 of Nocedal & Wright [26].

As in the previous chapter we consider methods for solving a general *constrained optimization problem*

$$\min_{\mathbf{x}\in\Omega} \quad f(\mathbf{x}) \tag{11.1}$$
$$\Omega = \{\mathbf{x} \in \mathbb{R}^n | c_i(\mathbf{x}) = 0,\ i \in \mathcal{E},\quad c_i(\mathbf{x}) \geq 0,\ i \in \mathcal{I}\}.$$

As the name implies, sequential quadratic programming (SQP) methods are iterative methods which solve at each iteration a quadratic programming problem (QP). For the problem (11.1) at the kth iteration (where we start from the current iterate $\mathbf{x}_k$) one solves for the next search direction

$$\min_{\mathbf{p}} \qquad \frac{1}{2}\mathbf{p}^T W_k \mathbf{p} + \boldsymbol{\nabla} f_k^T \mathbf{p} \tag{11.2a}$$
$$s.t. \qquad \boldsymbol{\nabla} c_i(\mathbf{x}_k)^T \mathbf{p} + c_i(\mathbf{x}_k) = \mathbf{0}, \quad i \in \mathcal{E} \tag{11.2b}$$
$$\boldsymbol{\nabla} c_i(\mathbf{x}_k)^T \mathbf{p} + c_i(\mathbf{x}_k) \geq \mathbf{0}, \quad i \in \mathcal{I}. \tag{11.2c}$$

For the subproblems (11.2) we can use the QP techniques described in Chapter 9. Here W_k is usually a positive semi-definite approximation of $\mathcal{L}_{xx}(\mathbf{x}_k, \boldsymbol{\lambda}_k)$. Most modern general-purpose software is based on SQP techniques: Consult the NEOS web site

`http://www-fp.mcs.anl.gov/otc/Guide/`

Let us restrict attention to the equality constrained case, i.e. assume for the moment that $\mathcal{I}$ is empty. (We can always use a gradient projection approach for inequality constraints as described in Section 9.3, so the equality-only case is important beyond its direct applicability.) Recall that our development of methods for unconstrained optimization has depended heavily on modeling the objective function f by a quadratic and solving nonlinear equations by linearizing them. This here is a direct extension of the same principles. Line search and trust region approaches can be devised. Indeed, almost everything that we have learned so far in this course comes into play in SQP methods!

Basic SQP and Newton-Lagrange

The SQP iteration for the equality constrained problem is now written as

$$\begin{aligned} \min_{\mathbf{p}} \quad & \frac{1}{2}\mathbf{p}^T W_k \mathbf{p} + \nabla f_k^T \mathbf{p} \\ s.t. \quad & \mathbf{c}_k - A_k \mathbf{p} = \mathbf{0}. \end{aligned} \tag{11.3}$$

(Recall Section 9.1 and also (9.19), (9.20).)[1] Assume that the constraint Jacobian, $-A_k$, has a full row rank m, and that $Z_k^T W_k Z_k$ is s.p.d., where Z_k is an $(n-m) \times n$ matrix of rank $n-m$ spanning the null-space of A_k: $A_k Z_k = 0$. Then $\mathbf{p}_k$ solves the QP (11.3) iff the KKT conditions

$$\begin{pmatrix} W_k & A_k^T \\ A_k & 0 \end{pmatrix} \begin{pmatrix} \mathbf{p}_k \\ \boldsymbol{\mu}_k \end{pmatrix} = \begin{pmatrix} -\nabla f_k \\ \mathbf{c}_k \end{pmatrix} \tag{11.4}$$

hold.

On the other hand, returning to the original problem (11.1) with no inequality constraints, the KKT conditions read

$$\begin{aligned} \nabla f(\mathbf{x}) + A(\mathbf{x})^T \boldsymbol{\lambda} &= \mathbf{0}, \\ \mathbf{c}(\mathbf{x}) &= \mathbf{0}. \end{aligned} \tag{11.5}$$

These are $n+m$ nonlinear equations in these many unknowns, and Newton's method for nonlinear equations (referred to as the *Newton-Lagrange method*) reads

$$\begin{pmatrix} \mathcal{L}_{\mathbf{xx}}(\mathbf{x}_k \boldsymbol{\lambda}_k) & A_k^T \\ A_k & 0 \end{pmatrix} \begin{pmatrix} \mathbf{p}_k \\ \delta\boldsymbol{\lambda}_k \end{pmatrix} = \begin{pmatrix} -\nabla f_k - A_k^T \boldsymbol{\lambda}_k \\ \mathbf{c}_k \end{pmatrix}. \tag{11.6}$$

Then $\mathbf{x}_{k+1} = \mathbf{x}_k + \mathbf{p}_k$, $\boldsymbol{\lambda}_{k+1} = \boldsymbol{\lambda}_k + \delta\boldsymbol{\lambda}_k$.

[1] Regrettably, there is a slight sign inconsistency with (10.1c) in that here

$$A_k^T = -\left[\nabla c_1(\mathbf{x}_k), \ldots, \nabla c_m(\mathbf{x}_k)\right].$$

Comparing (11.6) to (11.4) we see that the same method is obtained with $W = \mathcal{L}_{\mathbf{xx}}$ and $\boldsymbol{\mu}_k = \boldsymbol{\lambda}_{k+1}$. This gives the local SQP a familiar ring and immediately provides it with a local convergence theorem at a *quadratic* rate.

To incorporate inequality constraints we have already mentioned the gradient projection approach (which is applicable not only for QP). Another possibility is an active set method as described in Section 9.2. The solution from the previous iteration is used to start the method for the current iteration, thus providing also a good guess for the working set. Care has to be taken, however, to address infeasibility potentially brought in by such linearizations of inequality constraints.

This concludes the most simple minded application of SQP, as a locally convergent method. For a general-purpose implementation, however, several additional considerations arise.

Computing the search direction

Methods for the solution of KKT systems (11.4) have been discussed in Section 9.1. Let us recall the range-space method and the null-space method.

What is special in (11.4) is that close to convergence at least we expect $\|\mathbf{p}_k\| \to 0$, whereas $\boldsymbol{\mu}_k = \boldsymbol{\lambda}_{k+1}$ does not shrink likewise. Indeed, it is possible to find only $\mathbf{p}_k$, update $\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{p}_k$ (say), and then determine $\boldsymbol{\lambda}_{k+1}$ as if $\mathbf{p}_k = \mathbf{0}$, i.e. as the minimizer of the overdetermined least squares problem

$$\min_{\boldsymbol{\lambda}} \|A_k^T \boldsymbol{\lambda} + \boldsymbol{\nabla} f_k\|. \tag{11.7}$$

To recall, the null-space method finds $\mathbf{p}_k$ directly. A simplifying variant is referred to as a *reduced Hessian* method.

To maintain positive definiteness of the reduced Hessian and to avoid explicit specification of second derivatives, quasi-Newton methods may be used. There is a specially adapted BFGS for this purpose, see [26]. It is also possible to retain positive definiteness by using an augmented Lagrangian approach.

Merit function and descent

If one is to achieve a so-called global convergence (i.e., to expand the convergence basin of a Newton-Lagrange method) then line search or trust region techniques must be utilized. Note that these are utilized only for the primal variables, not the dual ones.

But then we need some way of deciding on what is a “sufficient decrease”. Indeed, what is a descent in the constrained context?! It cannot simply be any reduction in the objective value f in the presence of infeasibilities: the latter have to be taken into account as well.

Thus, a combination of penalty and barrier functions may be used as a *merit function*. One such useful merit function is an extension of (10.4),

$$\phi(\mathbf{x};\mu) = f(\mathbf{x}) + \frac{1}{\mu}\sum_{i\in\mathcal{E}} |c_i(\mathbf{x})| - \frac{1}{\mu}\sum_{i\in\mathcal{I}} \min\{0, c_i(\mathbf{x})\}. \tag{11.8}$$

Of course, now the question is, how should the parameter μ be chosen so that a coherent merit function results? It turns out that, considering only equality constraints, choosing at each iteration

$$\mu^{-1} > \|\boldsymbol{\lambda}_{k+1}\|_\infty \tag{11.9}$$

(μ is made not to change any more once the iterations apparently start to converge) yields that $\mathbf{p}_k$ is a descent direction with respect to ϕ of (11.8). So, line search using this merit function may commence.