Chapter 2

SQP Framework Algorithms

In this chapter, we outline two framework SQP algorithms. PDSQP requires the Hessian or Hessian approximation to be positive definite at every iteration, and consequently unique solutions to the QP subproblems are guaranteed. IDSQP does not require the Hessian or Hessian approximation to be positive definite at every iteration, so a unique (or even bounded) solution to the original QP subproblem is no longer guaranteed.

IDSQP is based on work by Murray and Prieto [MP95, MP99], who generalized their early-termination SQP algorithm ETSQP to incorporate exact second derivatives. The second-derivative algorithm, called SQP2D, combines a descent direction with a direction of negative curvature and performs a curvilinear search on the merit function at each major iteration. What interests us, in the context of this thesis, is that since the exact Hessians are not necessarily positive definite, Murray and Prieto developed a new strategy for finding a descent direction from a *nonconvex* QP subproblem. We present a simpler version of their second-derivative algorithm SQP2D that uses indefinite quasi-Newton approximations instead of exact second derivatives. We eliminate the use of directions of negative curvature from SQP2D and replace the curvilinear search with a standard linesearch. We call the remaining algorithm IDSQP, for "indefinite" or "indeterminate" SQP.

2.1 Positive-definite SQP algorithm (PDSQP)

PDSQP requires the Hessian or Hessian approximation to be positive definite at every iteration, and consequently unique solutions to the QP subproblems are guaranteed. How the

QP is solved is irrelevant to the convergence property of the SQP algorithm, but it is generally an iterative process with minor iterations (i.e., "QP iterations"). The quasi-Newton approximation to the Hessian of the Lagrangian is usually maintained using the BFGS quasi-Newton update to produce a positive-definite matrix B_k; this process is described in more detail in Chapter 3.

The main steps of each major iteration of algorithm PDSQP are:

- Solve the QP subproblem for a search direction $\ _k$ and multiplier estimate μ_k.
- Compute optimal slacks and slack search direction $\ _k$.
- Compute multiplier search direction ξ_k.
- Update penalty parameter ρ_k.
- Select steplength α_k using a linesearch on the merit function.
- Update iteration values $\ _{k+1}$, $\ _{k+1}$, $\ _{k+1}$, A_{k+1}, λ_{k+1}.
- Update Hessian approximation B_{k+1}.

2.1.1 Merit function and linesearch

What mainly distinguishes one positive-definite SQP algorithm from another is the choice of merit function used in the linesearch, which determines the steplength.

The SQP algorithms we consider in detail feature the *smooth augmented Lagrangian merit function*

$$L_A(\ \ \lambda\ \ \rho) = \ (\ ) - \lambda^T(\ (\ ) - \ ) + \frac{1}{2}\rho(\ (\ ) - \ )^T(\ (\ ) - \ ) \tag{2.1}$$

where $\rho > 0$ is the penalty parameter. The search space for this merit function includes the iterate and, in addition, the Lagrange multiplier estimates λ and the slack variables $\ \geq 0$. This merit function was suggested by Gill *et al.* [GMSW92] and a version of it has been successfully implemented in the dense SQP code NPSOL [GMSW86] and the large-scale SQP code SNOPT [GMS97]. A discussion of the smooth augmented Lagrangian function and other merit functions can be found in [Mur97].

Some notation

Components of the search direction on the subspaces corresponding to , λ, and will be denoted , ξ, and . For a fixed penalty parameter ρ, the value of the merit function ϕ as a function of the steplength α is denoted

$$\phi(\alpha; \quad \lambda \ \xi \quad \rho) \equiv L_A(\ + \alpha \quad \lambda + \alpha\xi \quad + \alpha \quad \rho). \tag{2.2}$$

We may abbreviate $\phi(\alpha; \ _k \ _k \ \lambda_k \ \xi_k \ _k \ _k \ \rho_k)$ by $\phi_k(\alpha)$ or by $\phi(\alpha)$ when the meaning is clear.

Slack variables

The slack variables appear within separable quadratic functions in the merit function; hence, their optimal values can be computed. At a given iterate, the optimal values of the slacks that minimize the merit function are

$$= \begin{cases} \max(0 \quad) & \text{if } \rho = 0 \\ \max(0 \quad - \lambda/\rho) & \text{otherwise.} \end{cases} \tag{2.3}$$

Search direction

The search direction comes directly from the QP subproblem. At each major iteration, the search direction in the space of slack variables is computed to satisfy

$$A \ - \ = -(\ - \ )$$

which ensures that the slacks remain at their optimal values for any linear constraints. The search direction in the space of the multiplier estimates is $\xi = \mu - \lambda$.

Derivative of the merit function

Consider the gradient of the merit function L_A with respect to , λ, and ,

$$\nabla L_A(\quad \lambda \quad) = \begin{pmatrix} (\ ) - \nabla \ (\ )^T\lambda + \rho\nabla \ (\ )^T(\ (\ ) - \ ) \\ -(\ (\ ) - \ ) \\ \lambda - \rho(\ (\ ) - \ ) \end{pmatrix}. \tag{2.4}$$

We may denote the derivative of ϕ_k with respect to α by ϕ_k'. It follows from (2.1) and (2.2) that $\phi'(0)$ is given by

$$\begin{aligned} \phi'(0) &= {}^T - {}^T A^T \lambda + \rho \; {}^T A^T(-) - (-)^T \xi + \lambda^T - \rho \; {}^T(-) \\ &= {}^T + (2\lambda - \mu)^T(-) - \rho \| - \|^2 \end{aligned}$$

where , $A = \nabla \; (\;)$, and are evaluated at .

Penalty parameter

The penalty parameter is adjusted as necessary at each iteration to ensure that (ξ) satisfies the desired descent condition

$$\phi'(0) \leq -\frac{1}{2} \; {}^T B \; . \tag{2.5}$$

If $\| - \| = 0$, it follows from the definition of the slack variables that ${}_0 = 0$ is an initial feasible point for the QP subproblem. Let the QP objective function be denoted $\psi(\;) \equiv {}^T + \frac{1}{2} {}^T H$. At the solution of the QP, $\psi(\;) = {}^T + \frac{1}{2} {}^T H \leq \psi({}_0) = 0$. Hence,

$$\phi'(0) = {}^T \leq -\frac{1}{2} \; {}^T H$$

implying that ρ does not need to be modified [MP95, pp. 603].

If $\| - \| > 0$, we obtain from (2.5) that for

$$\rho \geq \frac{({}^T + \frac{1}{2} {}^T B \;) + (2\lambda - \mu)^T(-)}{\| - \|^2} \tag{2.6}$$

we have

$$\phi'(0) = {}^T + (2\lambda - \mu)^T(-) - \rho \| - \|^2 \leq -\frac{1}{2} \; {}^T H$$

which implies the desired descent condition is satisfied.

In addition to satisfying (2.6), the penalty parameter must be adjusted in a manner that ensures it is not modified too often and maintains a minimum value of $\beta_\rho > 0$.

Algorithm PDSQP

$\leftarrow 0$
repeat
  $H \leftarrow B_k$
  Obtain the search direction $_k$ by calling PDQP to solve the QP subproblem
    $\min_p \quad {}_k^T \; + \frac{1}{2} \; {}^T H$
    s.t. $\quad A_k \; + \; {}_k \geq 0$
  Optimize the slack variables $_k$ (cf. (2.3))
  Form search direction for slacks $_k \leftarrow A_k \; {}_k + \; {}_k - \; {}_k$
  Form μ_k, an estimate of λ^* such that $\|\mu_k\| \leq \beta_\mu$
  Form multiplier search direction $\xi_k \leftarrow \mu_k - \lambda_k$
  if $\phi_k'(0) \leq -\frac{1}{2} \; {}_k^T B_k \; {}_k$
    $\rho_k \leftarrow \rho_{k-1}$
  else
    Increase ρ_k by at least a factor of 2, and
    large enough to satisfy (2.5) and (2.6)
  end
  Perform linesearch on the merit function to choose α_k (Figure 2.1).
  $\begin{pmatrix} {}_{k+1} \\ \lambda_{k+1} \end{pmatrix} \leftarrow \begin{pmatrix} {}_k \\ \lambda_k \end{pmatrix} + \alpha_k \begin{pmatrix} {}_k \\ \xi_k \end{pmatrix}$
  Evaluate $_{k+1}$, $_{k+1}$, and A_{k+1}
  Update B_k to form B_{k+1}
  $\leftarrow \; + 1$
until *convergence*

Figure 2.2: Algorithm PDSQP.

in Section 1.3.4:

BQP	$\underset{p}{\text{minimize}} \quad \psi(\ ) \equiv \ ^T \ + \frac{1}{2} \ ^T H$
	subject to $A \ = - \qquad \geq - \ .$

Notation

To avoid additional notation, there is an overlap of notation between the descriptions of the SQP algorithm and the QP algorithm, with , A, and Z having separate meanings in the context of the major (SQP) iterations and the minor (QP) iterations. To help avoid confusion, the subscript is used only to denote major iterations, while the subscript is used only in the QP iterations. Also note that in the minor iterations, the QP Hessian is H, while the Hessian approximation in the major iterations is denoted B.

The following notation and definitions will be used whenever we solve or compute a search direction from BQP. Suppose an initial feasible step $_0$ is given. Let A_0 be the matrix composed of the columns of A whose variables are free at $_0$, and let Z_0 be the matrix whose columns are a basis for the null space of A_0. Extending our notation, let A_j denote the matrix composed of the columns of A whose variables are free at the th iteration and let Z_j be the matrix whose columns span the null space of A_j. Note that we define "iteration" such that is incremented whenever there is a change to the active set. This includes both when a step is taken and $_j$ changes, as well as when a constraint is deleted from the active set but $_j$ does not change.

We introduce a permutation matrix

$$P_j \equiv \begin{pmatrix} \bar{P}_j & \widehat{P}_j \end{pmatrix} \tag{2.7}$$

such that the elements of $\ \equiv P_j^T \ $, where $_i = \ $, give the reordering of the variables by placing the indices of the free variables first followed by those of the fixed variables.

Define H_j to be the Hessian restricted to the free variables, $H_j \equiv \bar{P}_j^T H \bar{P}_j$. Let $_j = \bar{P}_j^T \ $. Note that $_j$ is the th iterate in the full variable space.

Algorithm PDQP for Positive-Definite Quadratic Programming

Obtain an initial feasible point $_0$
Identify initial working set and compute P_0, A_0, Z_0, and H_0
$\leftarrow 0$
repeat
  repeat
    $_j \leftarrow \bar{P}_j^T(\quad + H \quad _j)$
    stationary_point $\leftarrow Z_j^T \quad _j = 0$
    if not *stationary_point* **then**
      Solve for satisfying $\begin{pmatrix} H_j & A_j^T \\ A_j & 0 \end{pmatrix} \begin{pmatrix} \\ - \end{pmatrix} = \begin{pmatrix} - \quad _j \\ 0 \end{pmatrix}$
      $^- \leftarrow P_j^T \begin{pmatrix} \\ 0 \end{pmatrix}$
      $\gamma_M \leftarrow \min_\tau \left\{ \frac{_\tau + \quad _\tau}{| \, ^-_\tau|} \mid \, ^-_\tau < 0 \right\}$
      hit_constraint $\leftarrow \gamma_M < 1$
      $\gamma \leftarrow$ **if** *hit_constraint* **then** γ_M **else** 1
      $_{j+1} \leftarrow \quad _j + \gamma \, ^-$
      Update working set and compute P_{j+1}, A_{j+1}, Z_{j+1}, and H_{j+1}
      $\leftarrow \quad + 1$
    end
  until *stationary_point*
  Compute the Lagrange multipliers $\tilde{\sigma}_{FX}$ for the active bounds from (2.11)
  $\tilde{\sigma}_\tau \leftarrow \min_i \tilde{\sigma}_i$
  converged $\leftarrow \tilde{\sigma}_\tau \geq 0$
  if not *converged* **then**
    Delete bound with multiplier $\tilde{\sigma}_\tau$
    Update working set and compute P_{j+1}, A_{j+1}, Z_{j+1}, and H_{j+1}
    $\leftarrow \quad + 1$
  end
until *converged*

Figure 2.3: Algorithm PDQP.

Constrained stationary point

Moving to a constrained stationary point is a key component of algorithm PDQP. To explain this key component in slightly more general terms, we consider the Hessian of the BQP to be $H^{(0)}$ instead of H, where $H^{(0)}$ may be indefinite as long as any reduced Hessian encountered is positive definite. As before, let $H_j^{(0)}$ be the Hessian restricted to the free variables at iteration .

The necessary conditions that hold at a constrained stationary point $\tilde{}$ of BQP (1.17) are, for some multiplier vectors and $\tilde{\sigma}$:

$$\begin{aligned} &+ H^{(0)}\tilde{} = A^T \quad + \tilde{\sigma} \\ &A\tilde{} + \quad = 0 \\ &\tilde{} \geq - \\ &\tilde{\sigma}^T(\tilde{} + \quad) = 0. \end{aligned} \tag{2.8}$$

If at the stationary point we partition the variables into *fixed (FX)* and *free (FR)* using the permutation matrix P from (2.7), the Hessian can be written

$$\begin{aligned} H^{(0)} &= \begin{pmatrix} \bar{P}^T H^{(0)} \bar{P} & \bar{P}^T H^{(0)} \widehat{P} \\ \widehat{P}^T H^{(0)} \bar{P} & \widehat{P}^T H^{(0)} \widehat{P} \end{pmatrix} \\ &\equiv \begin{pmatrix} H_{FR}^{(0)} & H_{FRX}^{(0)T} \\ H_{FRX}^{(0)} & H_{FX}^{(0)} \end{pmatrix} \end{aligned} \tag{2.9}$$

and the submatrices H_{FX}, H_{FR}, and H_{FRX} are defined accordingly.

In terms of the free and fixed variables, the constrained stationary point $(\tilde{}_{FR} \ \tilde{}_{FX})$, , and $\tilde{\sigma}$ then satisfy

$$\begin{pmatrix} A_{FR} & A_{FX} \\ 0 & I_{FX} \end{pmatrix} \begin{pmatrix} \tilde{}_{FR} \\ \tilde{}_{FX} \end{pmatrix} = \begin{pmatrix} - \\ - \ _{FX} \end{pmatrix} \tag{2.10}$$

$$\begin{pmatrix} A_{FR}^T & \\ A_{FX}^T & I_{FX} \end{pmatrix} \begin{pmatrix} \\ \tilde{\sigma} \end{pmatrix} = \begin{pmatrix} _{FR} \\ _{FX} \end{pmatrix} + \begin{pmatrix} H_{FR}^{(0)}\tilde{}_{FR} - H_{FRX}^{(0)T} \ _{FX} \\ H_{FRX}^{(0)}\tilde{}_{FR} - H_{FX}^{(0)} \ _{FX} \end{pmatrix}.$$

Step to the stationary point

The null-space active-set method computes the step to the stationary point by starting from an initial feasible point and then forming at every iteration an equality-constrained QP (EQP) in the current set of free variables. The solution to the EQP is used as a search direction for the active-set method. Suppose there are $\bar{}$ free variables. At iteration , the EQP (where represents the free variables only) is

$$\boxed{\begin{array}{lll} \text{EQP} & \underset{d \in I\!R^{\bar{n}}}{\text{minimize}} & {}_j^T \; + \frac{1}{2} \; {}^T H_j^{(0)} \\ & \text{subject to} & A_j \; = 0 \end{array}}$$

where ${}_j = \bar{P}_j^T(\; + H^{(0)} \; {}_j)$.

For a minimizer * of the EQP to exist, $Z_j^T H_j^{(0)} Z_j$ must be positive definite (but $H_j^{(0)}$ is not necessarily positive definite). The KKT system representing the necessary conditions for the minimizer of the EQP are

$$\begin{pmatrix} H_j^{(0)} & A_j^T \\ A_j & 0 \end{pmatrix} \begin{pmatrix} {}^* \\ - \end{pmatrix} = \begin{pmatrix} - \; {}_j \\ 0 \end{pmatrix}.$$

The solution to the EQP can also be represented in terms of the null-space matrix Z_j, as

$${}^* = Z_j \; {}^Z \qquad \text{where} \qquad Z_j^T H_j^{(0)} Z_j \; {}^Z = -Z_j^T \; {}_j.$$

After the direction * is computed, the iterate is incremented by a step taken, in that direction, of length $\gamma \leq 1$

$$\leftarrow \; + \gamma P^T \begin{pmatrix} {}^* \\ 0 \end{pmatrix}.$$

The steplength γ is chosen to maintain feasibility with respect to all bounds. Eventually either a vertex is reached or the unit step is taken, at which point the requirements for a stationary point (2.10) are satisfied at the current point $\tilde{} =$ and .

Releasing a variable from its bound

The multipliers $\tilde{\sigma}_{FX}$ for the active bounds are computed from

$$\tilde{\sigma}_{FX} \quad = \quad {}_{FX} + H^{(0)}_{FRX}\,\tilde{}_{FR} - H^{(0)}_{FX}\;{}_{FX} - A^{T}_{FX} \tag{2.11}$$

or in the extended multipliers,

$$\tilde{\sigma} \quad = \quad + H^{(0)}(\bar{P}_j\bar{P}_j^T)\,\tilde{} - H^{(0)}(\widehat{P}_j\widehat{P}_j^T) \quad - A^{T} \quad . \tag{2.12}$$

If one of the multipliers is negative, is incremented, the corresponding variable is released from its bound, and new values are computed for P_j, A_j, Z_j, and H_j. The algorithm terminates with the solution when $\tilde{\sigma}_{FX} \geq 0$.

2.1.4 Convergence results for positive-definite SQP algorithm

The convergence results we give for PDSQP are based on results by Murray and Prieto [MP95] for a related but more general algorithm. What makes their algorithm (called ETSQP) more general is that it allows a wider variety of multiplier estimates and allows an incomplete solution of the subproblem. ETSQP allows any constrained stationary point of the QP subproblem, when combined with a descent direction taken from the stationary point and formed using the negative multiplier estimates at the stationary point, to be a search direction for the QP subproblem. Their approach limits the amount of work required in the subproblems by allowing "early termination" (hence the name ETSQP).

Under certain assumptions, global convergence for PDSQP follows from Murray and Prieto's results for ETSQP. In particular, the sequence $\{\ _k\}$ generated by PDSQP converges to a unique KKT point and λ_k converges to λ^*. Moreover, under additional assumptions on the quality of the Hessian approximation, it can be shown that the penalty parameter is bounded and the rate of convergence is superlinear.

We list below the precise assumptions needed to prove these results. Assumption A5, that every subproblem has a feasible solution, may be satisfied by modifying the SQP algorithm to allow infeasible constraints (see [MP95, GMS97, Bom99]). With this caveat, the first set of general assumptions all relate to properties of the problem, not the iterates or the algorithm. Assumptions MC3 and HC3 are needed only to prove the superlinear rate of convergence, not for global convergence. Note, it is not necessary to assume the iterates lie in a compact region. See [MP95, MP99] for a more detailed discussion of these assumptions.

Except for assumption HC2 (that the Hessian approximation is positive definite), we shall continue to make these assumptions for all SQP algorithms discussed.

General assumptions

A1. For some constant $\beta_c > 0$, the global minimum of the problem

$$\begin{array}{ll} \underset{x \in I\!R^n}{\text{minimize}} & F(\ ) \\ \text{s.t.} & (\ ) \geq -\beta_c \end{array}$$

exists.

A2. There exist no KKT points at infinity for problem NP (we use the definition of "KKT point at infinity" given in [MP95]).

A3. $F(\ )$, $(\ )$ and their first and second derivatives are continuous and uniformly bounded in norm on a compact set.

A4. The Jacobian corresponding to the active constraints at all KKT points has full rank.

A5. A feasible point $\ _{0_k}$ exists to all the QP subproblems, satisfying

$$\|\ _{0_k}\| \leq \beta_{p0} \|\tilde{\ }^{-}_{k}\| \qquad \ ^T_k\ _{0_k} \leq \beta_{p0} \|\tilde{\ }^{-}_{k}\|$$

for some constant $\beta_{p0} > 0$, where $\tilde{\ }_k$ denotes the normalized constraints, $(\tilde{\ }_k)_i \equiv \ _k/(1 + \|(\ _k)_i\|)$, and $(\ _k)_i$ is the th row of A_k.

A6. Strict complementarity holds at all stationary points of NP, including stationary points at infinity, if they exist.

A7. The reduced Hessian of the Lagrangian is nonsingular at all KKT points.

Assumptions on the multiplier estimates

MC1. The estimates μ_k are uniformly bounded in norm; that is, $\|\mu_k\| \leq \beta_\mu < \infty$.

MC2. The complementarity condition $\mu_k^T(A_k\ _k + \ _k) = 0$ is satisfied at all major iterations.

Assumptions on the Hessian approximation

HC1. $\beta_{svB} < \infty$ is an upper bound on the largest eigenvalue of $\{B_k\}$.

HC2. $\beta_{lvB} > 0$ is a lower bound on the smallest eigenvalue of $\{B_k\}$.

Assumptions needed to prove superlinear rate of convergence

MC3. $\|\mu_k - \lambda^*\| = O(\| \ _k - \ ^*\|)$, where λ^* denotes the multiplier vector associated with a KKT point $\ ^*$ closest to $\ _k$.

HC3. Following Boggs, Tolle and Wang [BTW82], we assume

$$\|\widehat{Z}_j^T(B_k - \nabla_x^2\mathcal{L}(\ _k\ \lambda))\ _k\| = \ (\|\ _k\|)$$

where B_k is the approximation to $\nabla_x^2\mathcal{L}(\ _k\ \lambda)$, and $\widehat{Z}_k$ is a basis for the null space of $\widehat{A}_k$, the Jacobian at $\ _k$ of those constraints active at $\ ^*$, that is bounded in norm and has its smallest singular value bounded away from 0.

2.2 SQP algorithm with indefinite Hessian approximations

The *indefinite* SQP algorithm does not require the Hessian or Hessian approximation to be positive definite at every iteration, so a bounded *solution* to the original QP subproblem is no longer guaranteed, let alone a unique solution. Consequently, it does matter both how the QP is solved and how the initial feasible point is computed. Depending on convenience and availability, the QP Hessian may be formed from exact second derivatives or approximated by any general quasi-Newton update (for more detail refer to Chapter 4).

The indefinite and positive-definite algorithms (IDSQP and PDSQP) are similar in many ways. They both use the same merit function (2.1). The main differences have to do with the way of computing the search direction from the subproblem and the algorithm for adjusting the penalty parameter ρ to ensure a descent direction at every iteration.

As before, we consider the QP to have the form BQP (1.17), which was introduced in Section 1.3.4 and repeated in Section 2.1.3.

2.2.1 Computing a search direction from the indefinite QP subproblem

Murray and Prieto's strategy for computing a search direction from a nonconvex QP is based on modifying the QP Hessian at the initial feasible point so the initial reduced Hessian is positive definite. For the modified QP subproblem there exists a constrained stationary point, which can be found by the null-space active-set method. Combining the step to the stationary point with a certain direction of descent provides a satisfactory search direction for the merit function.

The main steps of the procedure IDQP to compute a search direction are:

- Find an initial feasible step $_0$ for the BQP.
- Modify the initial reduced Hessian to be positive definite.
- Compute the step to the (first) stationary point.
- Compute a descent step from the stationary point.

Initial feasible point

The initial point $_0$ must be feasible ($A\ _0 = -$ and $_0 \geq -$) and for some $\beta_{p0} > 0$ satisfy

$$\|\ _0\| \leq \beta_{p0}(\|\tilde{}\|^2 + \|\ ^-\|^2)^{1/2} \qquad ^T\ _0 \leq \beta_{p0}(\|\tilde{}\|^2 + \|\ ^-\|^2)^{1/2}. \tag{2.13}$$

When the minimizer of a positive-definite QP is used as the search direction, then since the solution is unique, the choice of $_0$ is irrelevant. If we determine the search direction from a stationary point that is not a minimizer, the sequence of stationary points that we compute depends directly on the value of $_0$. These conditions on $_0$ ensure that all stationary points are satisfactory points at which to terminate the solution process.

Modified reduced Hessian

If the initial reduced Hessian $Z_0^T H Z_0$ is not positive definite, H is modified. Let $H^{(0)}$ be a modification of H for which $Z_0^T H_0^{(0)} Z_0 \succ 0$. If $Z_0^T H Z_0$ can be formed, Murray and Prieto suggest determining $H^{(0)}$ using the method described in [FGM95]. Let $\psi^{(0)}$ be the corresponding QP objective

$$\psi^{(0)}(\ ) = \frac{1}{2}\ ^T H^{(0)}\ +\ ^T\ .$$

In the large-scale case, the issue is more difficult, since merely forming $Z_0^T H Z_0$ may be prohibitively expensive. Fortunately, the nature of this matrix may be deduced from a factorization of the KKT matrix

$$\begin{pmatrix} H & A_0^T \\ A_0 & 0 \end{pmatrix}.$$

Forsgren and Murray [FM93] describe how to determine $H^{(0)}$ using the KKT matrix.

Alternatives for how to compute $H^{(0)}$ in both the large-scale and the small-scale cases are described in more detail in Chapter 5.

Existence of constrained stationary point

We shall prove a constrained stationary point exists in the case where $H^{(0)}$ is indefinite.

Lemma 2.2.1 *Consider the problem BQP with objective $\psi^{(0)}(\ )$. If $Z_0^T H^{(0)} Z_0$ is positive definite, the active-set method starting from $_0$ will reach a constrained stationary point (2.8) in a finite number of iterations.*

Proof. The initial working set, the set of free and fixed variables, and the nullspace matrix Z_0 all depend on the initial feasible point $_0$. A consequence of $H^{(0)}$ being chosen such that $Z_0^T H_0^{(0)} Z_0 \succ 0$ is that the solution to the EQP on the initial fixed variables,

$$\boxed{\begin{array}{ll} \underset{p \in I\!R^n}{\text{minimize}} & {}^T \ + \frac{1}{2} \ {}^T H^{(0)} \\ \text{subject to} & A \ = 0 \qquad {}_{FX} = - \ {}_{FX} \end{array}}$$

is bounded below, hence a minimizer exists to this problem. If the minimizer satisfies $_{FR} \geq - \ _{FR}$, a constrained stationary point has been reached. Otherwise, a step towards the minimizer is taken, the working set changes and the problem is solved again.

From $_0$, bounds are not deleted from the working set until a minimizer is reached. Since the first EQP is bounded below, when additional bounds are added the EQP remains bounded, and the steps taken are in a subset of the initial free variables. This manner of changing the working set ensures the reduced Hessian is positive definite on every subspace encountered. The algorithm continues until the working set defines a unique point or the unit step is taken so a minimizer is reached. In either case we have reached a constrained stationary point in a finite number of iterations. ∎

Acceptable stationary stopping points

Under a certain stronger positivity assumption on the Hessian of the modified QP, the proof of convergence for Murray and Prieto's second-derivative algorithm SQP2D allows the algorithm to continue to a stationary point $\tilde{}$ past the first one. In such a case, the algorithm would proceed as in PDQP, where the bound corresponding to the minimum eigenvalue is deleted from the set of active bounds, the reduced Hessian is updated, etc., until the next stationary point is reached. We need to ensure that the reduced Hessian is positive definite on all subspaces encountered, otherwise a minimizer on the subspace would not exist.

Suppose the Hessian has been modified at most once, at the initial feasible point. Let $\tilde{}$ be a stationary point, and let $\tilde{H} \equiv H^{(0)}$ be the Hessian at $\tilde{}$.

The general *positivity assumption* made in [MP99] requires that the modified Hessian be sufficiently positive definite on the union of the nullspaces encountered when determining a stationary point. In other words, $V^T\tilde{H}V \succ 0$, where the columns of V are a basis for all null-spaces encountered. For example, this ensures that $(\tilde{} - {}_0)^T\tilde{H}(\tilde{} - {}_0) \succ 0$, where ${}_0$ is the initial feasible point and $\tilde{}$ is the final stationary point.

Descent step off stationary point

In SQP2D, when the stationary point $\tilde{}$ is not a minimizer, it is combined with a multiple $\hat{\gamma}$ of a sufficient descent step that moves off several of the constraints with negative multipliers.

In the discussion that follows, let $\tilde{}$ be a stationary point but not a minimizer of the QP. Let $\tilde{\sigma}$ be the multiplier estimates at $\tilde{}$ for the bounds and let $\tilde{}$ be the multiplier estimates at $\tilde{}$ for the general constraints.

Let $\tilde{H}$ be the Hessian at $\tilde{}$ (for algorithm IDQP, $\tilde{H} = H^{(0)}$, but in Chapter 5 it could vary). Let ${}_{FX}$ be the components of corresponding to fixed variables at $\tilde{}$, and ${}_{FR}$ be the components of corresponding to free variables. The descent step is required to be a direction of sufficient descent from $\tilde{}$ satisfying the following conditions:

DD1. The direction is feasible and of unit length:
$A \quad = 0, \quad {}_{FX} \geq 0$, and $\| \quad \|_\infty = 1$.

DD2. The rate of descent along is "sufficiently" large, i.e., for some $0 < \beta_u \leq 1$, satisfies

$$\tilde{}^T \quad \leq \beta_u \tilde{}^T \ {}^* \tag{2.14}$$

where $\tilde{} \equiv \tilde{H}\,\tilde{} +$ and * solves

$$\begin{array}{ll} \min_u & \tilde{}^T \\ \text{s.t.} & A \;\; = 0 \\ & {}_{FX} \geq 0 \\ & \|\;\;\|_\infty \leq 1. \end{array} \tag{2.15}$$

Note that (2.14) and (2.15) imply $\tilde{}^T \;\; \leq 0$.

Murray-Prieto descent direction

The following descent direction was suggested in [MP99]. Assuming A has a bounded condition number, let $(\hat{} \;\; \hat{} \;\; \hat{\sigma})$ be the solution of

$$\begin{aligned} \tilde{H}\hat{} - A^T\hat{} - \hat{\sigma} &= -\tilde{} \\ A\hat{} &= 0 \\ \hat{}_{FX} &= \tilde{\sigma}^-_{FX}. \end{aligned} \tag{2.16}$$

Since $\tilde{}$ is not a minimizer of the QP, $\min_i \tilde{\sigma}_i < 0$ (so $\tilde{\sigma}^-_{FX} \neq 0$). System (2.16) is the set of first-order necessary conditions for the EQP

$$\boxed{\begin{array}{ll} \underset{u}{\text{minimize}} & \frac{1}{2}\;{}^T\tilde{H} \;\; + \tilde{}^T \\ \text{subject to} & A \;\; = 0 \qquad {}_{FX} = \tilde{\sigma}^-_{FX}. \end{array}}$$

The following argument shows that the direction $\;\equiv \hat{}/\|\hat{}\|_\infty$ satisfies conditions **DD1** and **DD2**. First, the direction is well-defined because under the assumptions on A, $\|\hat{}\|_\infty$ is bounded. Clearly $\|\;\;\|_\infty = 1$, and the rest of **DD1** follows from (2.16). Since at the stationary point $\tilde{} = A^T\tilde{} + \tilde{\sigma}$,

$$\tilde{}^T \;\; = (\tilde{}^T A + \tilde{\sigma}^T) \;\; = \tilde{\sigma}^T \;\; = -\frac{\tilde{\sigma}^T\tilde{\sigma}}{\|\hat{}\|_\infty} \leq -\frac{(\min_i \tilde{\sigma}_i)^2}{\|\hat{}\|_\infty}. \tag{2.17}$$

Together with

$$\tilde{}^T\,{}^* = (\tilde{}^T A + \tilde{\sigma}^T)\,{}^* = \tilde{\sigma}^T\,{}^* = \tilde{\sigma}^T_{FX}\,{}^*_{FX} \geq \;\; \min_i(\tilde{\sigma}_{FX})_i \tag{2.18}$$

this implies

$$\tilde{\ }^T \quad \leq -\frac{(\min_i \tilde{\sigma}_i)^2}{\|\hat{\ }\|_\infty} \leq \left(\frac{\min_i \tilde{\sigma}_i}{\|\hat{\ }\|_\infty}\right) \tilde{\ }^T \ {}^*. \tag{2.19}$$

If not for the general constraints, the resulting direction $\hat{\ }$ would be precisely a steepest descent direction. One might ask why we select a descent direction that fixes one or more newly freed variables at their multipliers (see (2.16)), and in doing so select the steepest-descent direction in those variables, rather than compute the Newton step to the minimizer. The reason is because this ensures the updated reduced Hessian is still positive definite. In Section 5.4.4, we describe an alternative descent direction that is the Newton step to a minimizer of a problem with a modified Hessian. There we address the problem of the updated reduced Hessian becoming indefinite.

Steplength

After normalizing the descent step $\hat{\ }$ so $\quad \equiv \hat{\ }/\|\hat{\ }\|$, a steplength $\tilde{\gamma}$ is computed, such that

- $\tilde{\ } + \tilde{\gamma} \quad$ is a feasible point for the QP,
- if $\ {}^T\tilde{H} \quad > 0$,

$$\tilde{\gamma} < -\frac{(\quad + \tilde{H}\tilde{\ })^T}{{}^T\tilde{H}}$$

  that is, the steplength is less than

$$\arg\min_{\gamma} \ {}^T(\tilde{\ } + \gamma \ ) + \frac{1}{2}(\tilde{\ } + \gamma \ )^T\tilde{H}(\tilde{\ } + \gamma \ ).$$

- $\tilde{\gamma} < \gamma_M$ ($\tilde{\gamma}$ is bounded above), and
- $\|\tilde{\ } + \tilde{\gamma} \ \| \geq \|\tilde{\ }\|$ (to ensure the resulting direction is sufficiently large).

Summary

At each iteration of **IDSQP** an inner iteration is performed to compute the search direction from the indefinite QP subproblem. The algorithm endeavors to solve the QP subproblem using an active-set method but terminates early in order to satisfy the positivity condition. In Figures 2.4–2.6, we provide pseudocode summarizing the algorithm. It is broken into three parts:

- Main subroutine IDQP: Compute-search-direction-from-indefinite-QP

- Subroutine Move-to-stationary-point
- Subroutine Compute-descent-direction

We assume that positive constants β_{p0}, β_{de}, β_{HZ}, β_B and $\gamma_M > 1$ have been defined. The subscript refers to the QP (inner) iterations.

Algorithm IDQP: Compute-search-direction-from-indefinite-QP

Obtain a feasible point $_0$ satisfying (2.13)
Identify initial working set and compute P_0, A_0, Z_0, and H_0
if $\lambda_{\min}(Z_0^T H_0 Z_0) > \beta_{HZ}$, **then**
    $H_0^{(0)} \leftarrow H_0$
else select $H_0^{(0)}$ so that $\lambda_{\min}(Z_0^T H_0^{(0)} Z_0) > \beta_{HZ}$ and $\|H^{(0)}\| \le \beta_H$
end
$\leftarrow 0$

$\boxed{\text{Move-to-stationary-point}}$

$\tilde{} \leftarrow {}_j$
$\tilde{H} \leftarrow H^{(0)}$
Compute the Lagrange multipliers $\tilde{\sigma}_{FX}$ for the active bounds from (2.11)
if $\tilde{\sigma}_{FX} \ge 0$, **then**
    $\tilde{\gamma} \leftarrow 0, \quad \leftarrow 0$
else

    $\boxed{\text{Compute-descent-direction}}$

end

Figure 2.4: Algorithm IDQP.

2.2.2 Adjusting the penalty parameter

We return to a feature of the merit function in the outer SQP algorithm, the penalty parameter. Let $= \tilde{} + \tilde{\gamma}$. The penalty parameter is adjusted as necessary at each iteration to ensure (ξ) is a descent direction for the merit function, i.e.,

$$\phi'(0) \le -\omega \tag{2.20}$$

Subroutine Move-to-stationary-point

repeat

$\quad {}_j \leftarrow \bar{P}_j^T(\ + H^{(0)} \ {}_j)$

$\quad$ *stationary_point* $\leftarrow Z_j^T \ {}_j = 0$

$\quad$ **if not** *stationary_point* **then**

$\quad\quad$ Solve for satisfying $\begin{pmatrix} H_j^{(0)} & A_j^T \\ A_j & 0 \end{pmatrix} \begin{pmatrix} \\ - \end{pmatrix} = \begin{pmatrix} - \ {}_j \\ 0 \end{pmatrix}$

$\quad\quad \bar{} \leftarrow P_j^T \begin{pmatrix} \\ 0 \end{pmatrix}$

$\quad\quad \gamma_M \leftarrow \min_\tau \left\{ \dfrac{\ {}_\tau + \ {}_\tau}{|\bar{}_\tau|} \mid \bar{}_\tau < 0 \right\}$

$\quad\quad$ *hit_constraint* $\leftarrow \gamma_M < 1$

$\quad\quad \gamma \leftarrow$ **if** *hit_constraint* **then** γ_M **else** 1

$\quad\quad {}_{j+1} \leftarrow \ {}_j + \gamma \bar{}$

$\quad\quad$ Update working set and compute P_{j+1}, A_{j+1}, Z_{j+1}, and $H_{j+1}^{(0)}$

$\quad\quad \leftarrow \ + 1$

$\quad$ **end**

until *stationary_point*

Figure 2.5: Subroutine Move-to-stationary-point.

Subroutine Compute-descent-direction

Compute $\hat{}$ that satisfies: $\begin{pmatrix} \tilde{H} & A^T & I \\ A & 0 & 0 \\ \hat{P}^T & 0 & 0 \end{pmatrix} \begin{pmatrix} \hat{} \\ -\hat{} \\ -\hat{\sigma} \end{pmatrix} = \begin{pmatrix} -\tilde{} \\ 0 \\ \tilde{\sigma}_{\bar{F}X} \end{pmatrix}$

Let γ_1 denote the largest feasible step from $\tilde{}$ along $\ \equiv \hat{}/\|\hat{}\|$

if $-(\ + \tilde{H}\tilde{})^T \ < \gamma_1 \ {}^T\tilde{H} \ $, **then**

$\quad \tilde{\gamma} \leftarrow \min\Big(-\dfrac{(\ + \tilde{H}\tilde{})^T}{{}^T\tilde{H}} \ \gamma_M\Big)$

else

$\quad \tilde{\gamma} \leftarrow \min(\gamma_1 \ \gamma_M)$

end

if $\|\tilde{} + \tilde{\gamma} \ \| < \|\tilde{}\|$, **then** $\tilde{\gamma} \leftarrow 0$

if ${}^T \ > 0$, **then** $\tilde{\gamma} \leftarrow 0$

Figure 2.6: Subroutine Compute-descent-direction.

where

$$\phi'(0) = \ ^T \ + (2\lambda - \mu)^T(\ - \ ) - \rho\| \ - \ \|^2 \tag{2.21}$$

and

$$\omega = \frac{1}{2}\left((\tilde{\ } - \ _0)^T \tilde{H}(\tilde{\ } - \ _0) + \| \ - \ \|^2 - 2\tilde{\gamma} \ ^T \ \right). \tag{2.22}$$

If (2.20) does not hold with the penalty parameter from the previous iteration, ρ is modified to satisfy

$$\rho \geq \frac{\frac{1}{2}\omega + \ ^T \ + (2\lambda - \mu)^T(\ - \ )}{\| \ - \ \|^2}. \tag{2.23}$$

If $\| \ - \ \| = 0$, then $\ _0 = 0$ and $\ ^T\tilde{\ } + \frac{1}{2}\tilde{\ }^T\tilde{H}\tilde{\ } \leq 0$, implying

$$\phi'(0) = \ ^T\tilde{\ } + \tilde{\gamma} \ ^T \ \leq -\frac{1}{2}\tilde{\ }^T\tilde{H}\tilde{\ } + \tilde{\gamma} \ ^T \ = -\omega$$

so no adjustment to ρ is necessary.

2.2.3 Statement of the algorithm

Given $\ _0$, $\ _0$, A_0, $\ _0$, B_0, and λ_0, select $\rho_{-1} \geq 0$, $0 < \sigma < \frac{1}{2}$, $\frac{1}{2} < \eta < 1$, $\beta_c \geq \| \ ^-(\ _0)\|_\infty$, $\beta_\mu \geq \|\lambda_0\|$, $\beta_H > 1$ and $\beta_\rho > 0$. The algorithm IDSQP is presented in Figure 2.7.

2.2.4 Convergence results for the indefinite SQP algorithm

For their second-derivative algorithm SQP2D, Murray and Prieto prove global convergence to a KKT point satisfying the second-order optimality conditions, quadratic convergence of both $\ _k$ and λ_k, and boundedness of the penalty parameter [MP99] (see Theorems 3.3, 4.1, and 4.2).

Under weaker assumptions, weakened versions of these results carry over to IDSQP. In particular, for any indefinite quasi-Newton Hessian approximation satisfying **HC3**, the same proofs imply IDSQP has global convergence to a KKT point at a superlinear rate [MP99].

Assumptions

The proofs of convergence that apply to IDSQP permit the stationary point $\tilde{\ }$ to be any constrained stationary point encountered by the active-set method, not only the first stationary point, as long as "the reduced Hessians of the modified QP are sufficiently positive

Algorithm IDSQP

$\leftarrow 0$
repeat
  $H \leftarrow B_k$
  Obtain a feasible step $_0$, a QP stationary point $\tilde{}$,
    a modified Hessian $\tilde{H}$, a sufficient descent direction ,
    and a feasible steplength $\tilde{\gamma}$ in the direction from $\tilde{}$,
    by calling IDQP on the QP subproblem

$$\begin{array}{ll} \min_p & {}_k^T \; + \tfrac{1}{2} \; {}^T H \\ \text{s.t.} & A_k \; + \; {}_k \geq 0 \end{array}$$

  $\bar{B}_k \leftarrow \tilde{H}$
  $_k \leftarrow \tilde{} + \tilde{\gamma}$
  Optimize the slack variables $_k$ (cf. (2.3))
  Form search direction for slacks $_k \leftarrow A_k \; _k + \; _k - \; _k$
  Form μ_k, an estimate of λ^*
  Form multiplier search direction $\xi_k \leftarrow \mu_k - \lambda_k$
  $\omega \leftarrow \frac{1}{2}((\tilde{} - \; _0)^T \bar{B}_k(\tilde{} - \; _0) + \| \; _k - \; _k\|^2 - 2\tilde{\gamma} \; {}_k^T \;)$
  if $\phi_k'(0) \leq -\frac{1}{2}\omega$
    $\rho_k \leftarrow \rho_{k-1}$
  else

$$\rho_k \leftarrow \max\left(2\rho_{k-1} \quad \frac{\frac{1}{2}\omega + \; {}_k^T \; _k + (2\lambda_k - \mu_k)^T(\; _k - \; _k)}{\| \; _k - \; _k\|^2} \quad \beta_\rho\right)$$

  end
  Perform a linesearch on the merit function to choose α_k (Figure 2.1).

$$\begin{pmatrix} _{k+1} \\ \lambda_{k+1} \end{pmatrix} \leftarrow \begin{pmatrix} _k \\ \lambda_k \end{pmatrix} + \alpha_k \begin{pmatrix} _k \\ \xi_k \end{pmatrix}$$

  Evaluate $_{k+1}$, A_{k+1} and $_{k+1}$
  Update B_k to form B_{k+1}
  $\leftarrow \; + 1$
until *converged*

Figure 2.7: Algorithm IDSQP.