TMA 4180 Optimeringsteori
Quadratic Programming Basics

H. E. Krogstad, NTNU

Spring 2005/Rev. 2008

Quadratic Programming (QP) is the next step up from Linear Programming. This material is somewhat technical and not very well suited to lecture on the blackboard. However, many problems may be formulated as QP problems, and QP constitutes the iteration step in the highly efficient and quite popular *Sequential Quadratic Programming* (SQP) methods. A certain basic knowledge of QP problems is therefore valuable.

The material in this note is also found in N&W, Chapter 16, and covers what we consider to be the curriculum for the present course.

Some of the arguments are shortened, and the note does not treat the numerical aspects in any great detail.

There is also an accompanying set of transparencies covering QP which will be used in the lectures.

1 The QP Problem

We are now considering problems where the objective function is *quadratic,*

$$q(x) = \frac{1}{2}x'Gx + d'x, \tag{1}$$

and where G is a symmetric $n \times n$ matrix. We recall that

$$\nabla q(x)' = Gx + d, \tag{2}$$

and if there are no constraints and $G > 0$, the unique solution is of course $x^* = -G^{-1}d$.

However, in the present case, we first of all *have* constraints, and, secondly, G may *not be positive definite,* not even *semi-definite*. Nevertheless, since q is continuous, we always have solutions as long as the feasible domain is bounded, see Fig. 1

Exactly as for LP problems, the *feasibility domain,* Ω, is also in this case defined in terms of *linear equality constraints*,

$$a_i'x = b_i,\ i \in \mathcal{E}, \tag{3}$$

and *linear inequality constraints*,

$$a_i'x \geq b_i,\ i \in \mathcal{I}. \tag{4}$$

Like in the LP case, Ω is convex, and q will be convex as long as $G \geq 0$ (which may not be the case).

Locally, even non-linear smooth constraints will look linear, e.g.

$$0 = c(x) \approx c(x_0) + \nabla c(x_0)(x - x_0). \tag{5}$$

Similarly, the quadratic form $q(x)$ could be a local approximation of a more general objective function. Thus, analyzing QP problems is useful also from a more general viewpoint.

Inequality constraints turn out to be somewhat troublesome, so we shall start with the simpler case of equality constraints only.

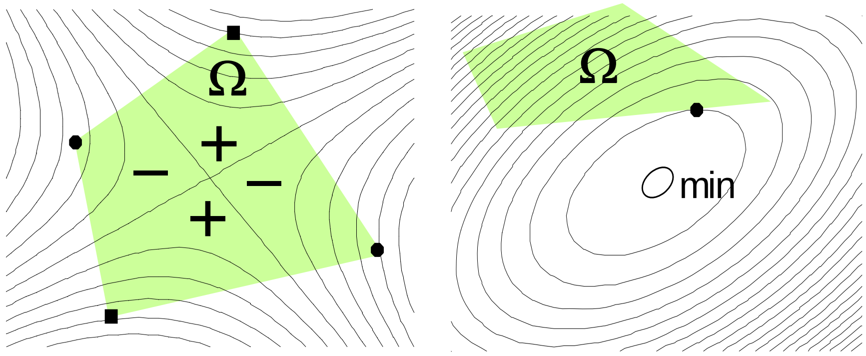

Figure 1: *Left*: If the G matrix is indefinite, there may be isolated local minima (circles) and maxima (squares). *Right*: If $G > 0$, both q and Ω are convex, and we have a (unique) minimum.

2 QP Problems with Equality Constraints

The case with only equality constraints can always be reduced to the following:

$$\begin{gathered}\min q(x) \\ Ax = b,\ A \text{ has full rank } r < n.\end{gathered} \tag{6}$$

(The reduction to a linear equation with full row rank is similar to the LP case). Let Z be a matrix consisting of basis vectors for the null-space, $\mathcal{N}(A)$, of A. We leave to the reader to prove that Ω, if non-empty, may always be written as:

$$\Omega = \{x;\ Ax = b\} = \left\{x_0 + Zu;\ Ax_0 = b,\quad u \in \mathbb{R}^{n-r}\right\}. \tag{7}$$

One way to proceed is therefore obvious: First find an x_0 so that $Ax_0 = b$ and determine a basis for $\mathcal{N}(A)$. Then insert $x = x_0 + Zu$ into $q(x)$ and obtain an *unconstrained* problem in u on $\mathbb{R}^{n-r}$:

$$\begin{aligned} f(u) &= q(x_0 + Zu) \\ &= \frac{1}{2}(x_0 + Zu)' G (x_0 + Zu) + d'(x_0 + Zu) \qquad (8) \\ &= \frac{1}{2} u' \tilde{G} u + \tilde{d}' u + \text{const.}, \qquad (9) \end{aligned}$$

where

$$\begin{aligned} \tilde{G} &= Z' G Z, \\ \tilde{d} &= Z'(Gx_0 + d). \end{aligned} \tag{10}$$

The only candidates for minima of the unconstrained problem

$$\min_{u \in \mathbb{R}^{n-r}} f(u) \tag{11}$$

are the solutions u^* such that

$$\tilde{G} u^* = -\tilde{d}, \tag{12}$$

which result in the following three cases:

1. A *unique solution* if $\tilde{G} > 0$,
2. *Infinitely many solutions* if $\tilde{G}$ is singular, as long as it is positive semi-definite and $\tilde{d} \in \mathcal{R}\left(\tilde{G}\right)$,
3. *No solution* if $\tilde{G}$ is not positive semi-definite, or $\tilde{d} \notin \mathcal{R}\left(\tilde{G}\right)$.

It is even possible to solve the KKT-equations for the problem in 6 directly. The Lagrange function is

$$\mathcal{L}(x, \lambda) = \frac{1}{2} x' G x + d' x - \lambda' (b - Ax), \tag{13}$$

and the KKT equations are simply

$$\nabla_x \mathcal{L}(x, \lambda)' = Gx + d + A'\lambda = 0, \tag{14}$$
$$Ax = b. \tag{15}$$

These equations may be collected into the combined system

$$\begin{bmatrix} G & A' \\ A & 0 \end{bmatrix} \begin{bmatrix} x \\ \lambda \end{bmatrix} = \begin{bmatrix} -d \\ b \end{bmatrix}. \tag{16}$$

The coefficient matrix is non-singular if A has full rank and G is positive definite on $\mathcal{N}(A)$ (Try to prove this yourself before you look at N&W Lemma 16.1, or the lemma stated on p. 17 in the KKT-note). If this is the case, we therefore have a unique solution which also is a global minimum. This is in accordance with point 1 above, since G is positive definite on $\mathcal{N}(A)$ if and only if $\tilde{G} = Z'GZ > 0$.

The solution for λ^* is obtained from Eqn. 14 if we know x^*:

$$\lambda^* = -\left(AA'\right)^{-1} A\left(d + Gx^*\right),$$

Numerical algorithms and an explicit formula for the inverse of the coefficient matrix in Eqn. 16 are described in N&W, Sec. 16.2 and 16.3.

3 QP Problems with Inequality Constraints

Inequality constraints spoil the elegant theory above completely!

If we have constraints of the familiar form

$$Ax = b,$$
$$x \geq 0, \tag{17}$$

one could try to use an elimination as we did for the LP case,

$$[B \ N] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = b. \tag{18}$$

However, this just leads to

$$\left(B^{-1}(b - Nx_2)\right)_i \geq 0, \ i \in \mathcal{I}_1,$$
$$(x_2)_i \geq 0, \ i \in \mathcal{I}_2, \tag{19}$$

that is, the same number of inequalities as before, and these are now much more complicated!

Let us then consider the general problem

$$\min_x \left\{ \frac{1}{2} x' G x + d' x \right\},$$
$$a_i' x - b_i = 0,\ i \in \mathcal{E}, \tag{20}$$
$$a_i' x - b_i \geq 0,\ i \in \mathcal{I}.$$

We assume that all equality constraints are linearly independent and let $\mathcal{A}$ or $\mathcal{A}(x)$ denote the active constraints at the point x. This is called the *active set* at x.

Show that the KKT-conditions are

$$Gx + d - \sum_{i \in \mathcal{A}(x)} \lambda_i a_i = 0, \tag{21}$$
$$a_i' x = b_i,\ i \in \mathcal{A}(x), \tag{22}$$
$$a_i' x > b_i,\ i \in \mathcal{I} \backslash \mathcal{A}(x), \tag{23}$$
$$\lambda_i \geq 0,\ i \in \mathcal{I} \cap \mathcal{A}(x), \tag{24}$$
$$\lambda_i \left(a_i' x - b_i \right) = 0,\ i \in \mathcal{I} \cup \mathcal{E}. \tag{25}$$

Recall that the LICQ condition is not necessary for linear constraints.

If x_0 is a KKT-point for the full problem, it is also a KKT-point for the reduced problem

$$\min_x \left\{ \frac{1}{2} x' G x + d' x \right\},$$
$$a_i' x = b_i,\ i \in \mathcal{A}(x_0). \tag{26}$$

(Proof left to the reader!). The reduced problem is a QP problem with equality constraints.

If we have an active set $\mathcal{A}$ consisting of $\mathcal{E}$ and some of $\mathcal{I}$, and have found a KKT-point x_0 for the reduced problem, it is easy to check the full KKT-conditions in Eqns. 21–24. If these are fulfilled, we have a KKT-point for the full problem. We need to find out whether this is a minimum, and the next step could therefore be to check the *2nd order* conditions: Form the matrix A consisting of the gradients of the active constraints and investigate whether $Z'GZ$ is positive (semi)definite, where Z is a basis for $\mathcal{N}(A)$.

The good news is therefore that if we know $\mathcal{A}$, we will be able to find a solution. The not so good news is that the number of combinations of active constraints grows exponentially with the number of constraints!

4 Active Set Methods

N&W contains a rather extensive treatment of iterative methods that try to adjust the active set towards the solution. We shall only give a brief discussion here, also assuming that $G \geq 0$ so that we have a *convex QP problem* with all solutions (if they exist) collected in a convex set.

Assume we are at a point $x_0 \in \Omega$. We then choose a *working set* $\mathcal{W}$ so that

$$\mathcal{E} \subset \mathcal{W} \subset \mathcal{A}(x_0). \tag{27}$$

(We always need to include $\mathcal{E}$ in $\mathcal{W}$, if not, we will not be in Ω!). Let $A_{\mathcal{W}}$ be the corresponding matrix of gradients of the constraints and solve the equality constrained (reduced) QP problem

$$\begin{aligned} &\min q\left(x_0+p\right) \\ A_{\mathcal{W}}&\left(x_0+p\right)=b_{\mathcal{W}}. \end{aligned} \tag{28}$$

If p turns out to be 0, we have to check whether x_0 could be the full solution. If not, we need to change the working set in some way (see below).

If $p \neq 0$, we consider p as a search direction and determine $\alpha \leq 1$ as the maximum value where

$$x_1 = x_0 + \alpha p \in \Omega. \tag{29}$$

If $\alpha = 1$, we are at a KKT-point for the reduced problem (the solution of 28), and since x_1 is in Ω, we also need to check the KKT-equations for the full problem. Otherwise, new inequality constraints have become active, which we now include in $\mathcal{W}$ and continue as above from x_1.

Sooner or later we end up in a case where $p = 0$ at a point x^* and with an active set $\mathcal{W}^*$. This point satisfies

$$Gx^* + d - \sum_{i\in\mathcal{W}^*} \lambda_i^* a_i = 0. \tag{30}$$

Let us define the Lagrange multipliers for the (inequality) constraints that are not in $\mathcal{W}^*$ to be 0. If now

$$\lambda_i^* \geq 0 \tag{31}$$

for all $i \in \mathcal{W}^* \cap \mathcal{I}$, we have reached a KKT-point for the full solution (check Eqns. 21 – 24!).

However, if some of these multipliers are negative, we throw the corresponding constraints out from $\mathcal{W}^*$ and solve a new reduced problem in the form of Eqn. 28.

It may be shown (Theorems 16.5 and 16.6 in N&W) that this will decrease the objective further.

In order to start the method, that is, to identify a feasible point $x_0 \in \Omega$, it may be necessary to carry out a Phase 1 problem as in the LP case.

The Active Set algorithm is listed on p. 472 in N&W, and numerical aspects are given on pp. 477 – 480. Each iteration involves minimization over a set

$$\left\{x_0 + Zu\right\}, \tag{32}$$

where Z contains a basis for the null-space of the matrix of gradients of active constraints. Since $\mathcal{W}$ only changes with a few constraints from iteration to iteration, the update of Z may be done numerically efficient.

QP for indefinite problems will not be discussed here, and is only very briefly discussed in the 2nd Ed. of N&W.

4.1 Example

Figure 2 illustrates an example from N&W, pp. 474 – 476 where $G > 0$. There are 5 inequality constraints denoted by A, B, $\cdots$, E in the figure. The red gradient arrows show the direction into the positive domains, and the constraints define a bounded, feasible domain Ω.

1. We start at the point $x^0 = (2, 0)$ with $\mathcal{W} = \{A, B\}$. Now, x^0 is a KKT-point for the corresponding reduced problem, but this brings us nowhere, since the feasible set for the reduced problem just contains x^0 ($p = 0$)!

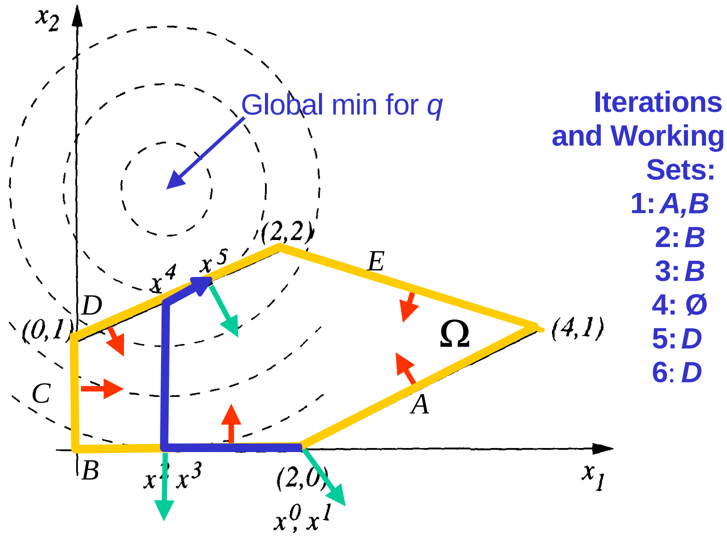

Figure 2: Active Set example. See text for explanation.

2. We choose in the next step to abandon the constraint associated with the most negative Lagrange multiplier (here this is A) and move along B ($\mathcal{W} = \{B\}$). This brings us from $x^1 = (2, 0)$ to x^2 (A *full* step, $\alpha = 1$).

3. This point is a KKT point for the new reduced problem, but *not* for the full problem.

4. No change is possible if we keep B in $\mathcal{W}$, so we abandon this as well and move from $x^3 = x^2$ in the negative gradient direction with $\mathcal{W} = \varnothing$. However, we can't continue up to $\alpha = 1$ since we, on our way to the unconstrained global minimum for q, meet D and have to stop at x^4.

5. Finally, we include D in $\mathcal{W}$ and continue to x^5, which turns out to be the solution x^*.

There are different roads towards the solution: What happens if we abandon *both* constraints in x^1?

5 The Gradient Projection Methods

The traditional gradient projection method admits non-linear objective functions as long as the constraints are linear:

$$
\begin{aligned}
&\min f(x), \\
&a_i' x = b_i, \ i \in \mathcal{E}, \\
&a_i' x \geq b_i, \ i \in \mathcal{I}.
\end{aligned} \tag{33}
$$

Assume that we are in a point $x_k \in \Omega$ with a corresponding set of active constraints $\mathcal{A}_k$ and the (full rank) matrix of gradients A_k. A *local feasible domain* is then defined as

$$
\Omega_k = \{x \ ; \ A_k x = b_k\} = \{x_k + \mathcal{N}(A_k)\}. \tag{34}
$$

Moving around in Ω_k does not violate Ω as long as we do not encounter new constraints. The gradient in x_k is

$$g_k = \nabla f(x_k)', \tag{35}$$

but in general $x_k - \alpha g_k$ will not be in Ω_k for any $\alpha \neq 0$. We therefore *project* the gradient onto $\mathcal{N}(A_k)$ and consider the 1-D problem

$$\min_{\alpha} f\left(x_k - \alpha P_{\mathcal{N}(A_k)} g_k\right), \tag{36}$$

$$x_k - \alpha P_{\mathcal{N}(A_k)} g_k \in \Omega. \tag{37}$$

Note that the projection is the vector in $\mathcal{N}(A_k)$ closest to g_k (Prove that the projection, $g_k \to P_{\mathcal{N}(A_k)} g_k \in \mathcal{N}(A_k)$ is defined by the *projection operator*

$$P_{\mathcal{N}(A_k)} = I - A_k' \left(A_k A_k'\right)^{-1} A_k, \tag{38}$$

by solving the equality constrained QP-problem

$$\begin{aligned} &\min_{x} \|g - x\|^2, \\ &Ax = 0.) \end{aligned} \tag{39}$$

The advantage of this approach is that many constraints may be changed for each iteration, but the disadvantage is the computations of the projection operators.

N&W describes an interesting variant which works for simple bound constraints in the QP-case (as far as I see, one can use the method for any non-linear objective function as long as the constraints are simple bounds):

$$\begin{aligned} &\min q(x), \\ &l \leq x \leq u. \end{aligned} \tag{40}$$

The bounds are to be understood component-wise, and include, if necessary, $-\infty$ and ∞.

Consider the following (trivial!) *non-linear* projection operator P_{lu} defined by

$$P_{lu}(x)_i = \begin{cases} l_i, & x_i \leq l_i \\ x_i, & l_i < x_i < u_i \\ u_i, & u_i \leq x_i \end{cases} \tag{41}$$

We start at x_0 and compute the continuous broken line path

$$x_{lu}(t) = P_{lu}\left(x_0 - t\nabla q(x_0)'\right), \tag{42}$$

as shown in Fig. 3. Of course, all $x(t) \in \Omega$, but the path will sooner or later stop. Let x_c be the first local minimum along the path,

$$x_c = \arg\min_{x(t)} f(x(t)).$$

From this point, the simplest would be to just compute a new gradient and repeat the operation. It is also possible to possible to do an approximate Active Set iteration using the already active bounds as the active set.

This latest version of the gradient projection algorithm turns out to have a wider applicability than one should expect. In fact, the so-called *Wolfe's Dual* of a convex QP problem, treated below, may *always* be formulated as a simple bound-constrained problem.

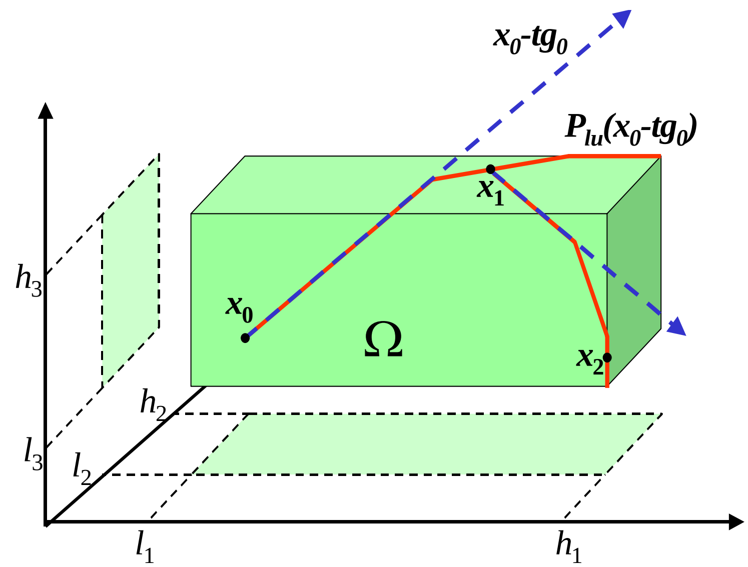

Figure 3: The non-linear gradient projection method where one searches for a minimum along the continuous broken line in each iteration.

6 Wolfe's Duality for Convex Problems

Before we start, it is useful to recall the KKT Theorem for convex problems, as discussed in the KKT-note. First of all, if

$$\Omega = \{x;\ c_i(x) = 0,\ i \in \mathcal{E},\ c_i(x) \geq 0,\ i \in \mathcal{I}\}, \tag{43}$$

where $c_i(x)$ for $i \in \mathcal{E}$ are linear functions, and $c_i(x)$ for $i \in \mathcal{I}$ are concave, then Ω is convex. We could, alternatively, require that all constraints written as inequalities are concave. This would imply that equality constraints had to be linear. If, in addition, the objective function $f(x)$ is convex, the *Convex KKT Theorem* then ensures that all KKT-points are global minima. The KKT-equations take the form

$$\begin{array}{ll} (i) & \nabla f(x^*) = \sum_{i\in\mathcal{I}} \lambda_i^* \nabla c_i(x^*), \\ (ii) & \lambda_i^* c_i(x^*) = 0, \\ (iii) & c_i(x^*) \geq 0, \\ (iv) & \lambda_i^* \geq 0. \end{array} \tag{44}$$

In linear programming the *Dual Problem* entered in a natural way by letting x and λ switch places, whereas the KKT-equations remained the same. Unfortunately, the situation is more complex for non-linear problems, but for convex problems there is a simple version of duality called the *Wolfe's Dual Problem*. In N&W 1st Ed. this was the problem that appears from nowhere in Section 16.8 (Eqn. 16.56a). In N&W 2nd Ed. it is hardly mentioned.

We now consider a convex problem where the Lagrangian is written

$$\mathcal{L}(x,\lambda) = f(x) - \lambda' c(x) = f(x) + \lambda' (-c(x)), \tag{45}$$

$$c(x) = (c_1(x), c_2(x), \cdots, c_m(x))', \tag{46}$$

and all constraints are rewritten as inequality constraints. Since we are really interested *only* in cases where $x \in \Omega$ and $\lambda \geq 0$, $\mathcal{L}$ will, as in the proof of the Convex KKT Theorem, be a sum of convex functions with positive weights, and therefore itself be convex in x.

After this introduction, it is possible to consider a simple version of duality, namely *Wolfe's Dual Problem*:

$$\begin{aligned} \max_{x,\lambda} \mathcal{L}(x, \lambda), \\ \nabla_x \mathcal{L}(x, \lambda) = 0, \\ \lambda \geq 0. \end{aligned} \tag{47}$$

Note that for this problem, *all* constraints for x are contained in the first equality constraints, and that all constraints for λ are linear.

Theorem: *Assume that f and c_i are differentiable and that x^* is a KKT point (and hence solves the primal problem). Then x^* and λ^* from Eqns.* 44 *will solve Wolfe's Dual Problem. Moreover,*

$$f(x^*) = \mathcal{L}(x^*, \lambda^*). \tag{48}$$

Proof: Since the KKT-equations guarantee that

$$\nabla_x \mathcal{L}(x^*, \lambda^*) = \nabla f(x^*) - \nabla c(x^*)\lambda^* = 0 \tag{49}$$

for some $\lambda^* \geq 0$, the pair (x^*, λ^*) will be a feasible point for problem (47). Since the KKT-equations also imply that

$$\lambda_i^* c_i(x^*) = 0, \ i = 1, \cdots, m, \tag{50}$$

this in turn shows that $f(x^*) = \mathcal{L}(x^*, \lambda^*)$.

If (x, λ) is another feasible pair for Eqn. 47, the Lagrangian will still be convex in x since $\lambda \geq 0$, and, as in the proof of the Convex KKT Theorem,

$$\begin{aligned} \mathcal{L}(x^*, \lambda^*) &= f(x^*) \\ &\geq f(x^*) - \sum_{i \in \mathcal{I}} \lambda_i c_i(x^*) \\ &= \mathcal{L}(x^*, \lambda) \\ &\geq \mathcal{L}(x, \lambda) + \nabla_x \mathcal{L}(x, \lambda) \cdot (x^* - x) \\ &= \mathcal{L}(x, \lambda). \end{aligned} \tag{51}$$

(Recall that $\nabla_x \mathcal{L}(x, \lambda) = 0$ is the feasibility requirement on (x, λ) in Eqn. 47).

We leave to the readers to prove that Wolfe's dual is equivalent to the regular dual problem for linear programming problems. However, for more general cases it does not hold that the "dual of the dual is the primal" since Wolfe's Dual Problem is not necessary convex.

Similarly to linear programming, it may be computationally advantageous to solve the dual problem, and Quadratic Optimization with inequality constraints is such a case. The Wolfe's Dual problem to the primal

$$\begin{aligned} \min_x \frac{1}{2} x' G x + d' x, \\ Ax \geq b, \end{aligned} \tag{52}$$

is

$$\max_{x,\lambda} \frac{1}{2}x'Gx + d'x - \lambda'(Ax - b),$$
$$Gx + d - A'\lambda = 0, \tag{53}$$
$$\lambda \geq 0$$

(check it!). When G is non-singular, we may solve for x right away,

$$x = G^{-1}\left(A'\lambda - d\right), \tag{54}$$

which in turn leads to a quadratic problem in λ:

$$\max_{\lambda} \left\{-\frac{1}{2}\lambda'\left(AG^{-1}A'\right)\lambda + \lambda'\left(b + AG^{-1}d\right) - \frac{1}{2}d'G^{-1}d\right\}, \tag{55}$$
$$\lambda \geq 0. \tag{56}$$

Note that the dimension of this problem is equal to the number of inequalities in $Ax \geq b$, and can be much less than the dimension of x. Since the constraints are now simply Eqn. 56, and the gradient of the objective function is

$$-\left(AG^{-1}A'\right)\lambda + b + AG^{-1}d,$$

the problem may be easily be attacked by the non-linear gradient-projection method.

Finally, x^* is found from

$$Gx^* = A'\lambda^* - d. \tag{57}$$

The material for this last section is adapted from the book of Fletcher: *Practical Methods of Optimization, 2nd Edition*, Wiley, 1996.

Duality for non-linear problems is briefly discussed in N&W, 2nd Ed., Sec. 12.9.