Projection Methods for Linear Equality Constrained Problems

Robert M. Freund

March, 2004

1 Review of Steepest Descent

Suppose we want to solve

$$P: \quad \text{minimize} \quad f(x)$$

$$\text{s.t.} \quad x \in \Re^n,$$

where $f(x)$ is differentiable. At the point $x = \bar{x}$, $f(x)$ can be approximated by its linear expansion

$$f(\bar{x} + d) \approx f(\bar{x}) + \nabla f(\bar{x})^T d$$

for d "small." This leads to the choice of d dictated by the direction-finding problem:

$$\text{minimize} \quad \nabla f(\bar{x})^T d$$

$$\text{s.t.} \quad \|d\| \leq 1,$$

which is equivalent to:

$$\text{minimize} \quad \nabla f(\bar{x})^T d$$

$$\text{s.t.} \quad d^T I d \leq 1.$$

The solution to this direction finding problem is:

$$\bar{d} = \frac{-\nabla f(\bar{x})}{\|\nabla f(\bar{x})\|}.$$

Because we choose our next step as

$$x' = \bar{x} + \alpha \bar{d}$$

for some choice of step-length α, then we can re-scale the direction $\bar{d}$ simply as:

$$\bar{d} = -\nabla f(\bar{x}).$$

That is, the steepest descent direction is simply the negative of the gradient of $f(x)$ at $x = \bar{x}$.

2 Equality Constrained Problems

Now consider the slightly more complicated problem

$$P: \quad \text{minimize} \quad f(x)$$

$$\text{s.t.} \quad Ax = b$$
$$x \in \Re^n,$$

where $f(x)$ is differentiable. The KKT conditions for this problem are as follows:

$$\begin{array}{rcl} A\bar{x} & = & b \\ \nabla f(\bar{x}) + A^T \bar{\pi} & = & 0. \end{array}$$

We wish to find such a KKT point.

Suppose that we are at the point $x = \bar{x}$, where $A\bar{x} = b$, i.e., $\bar{x}$ is a feasible point. Again we have

$$f(\bar{x} + d) \approx f(\bar{x}) + \nabla f(\bar{x})^T d$$

for d "small." In order to choose the direction $\bar{d}$ and compute the next point

$$x' = \bar{x} + \alpha \bar{d}$$

for some stepsize α, we will solve the following direction-finding problem:

$$\text{minimize} \quad f(\bar{x}) + \nabla f(\bar{x})^T (x - \bar{x})$$

$$\text{s.t.} \qquad Ax = b$$
$$\|x - \bar{x}\| \le 1,$$

or equivalently (by setting $d = x - \bar{x}$)

$$\text{minimize} \quad \nabla f(\bar{x})^T d$$

$$\text{s.t.} \qquad Ad = 0$$
$$d^T I d \le 1.$$

Note that $Ad = 0$ ensures that $A(\bar{x} + \alpha d) = A\bar{x} = b$ for any α. Also note that the constraint "$d^T I d \leq 1$" says that d must lie in the Euclidean unit ball B, defined as:

$$B = \{d \in \Re^n \mid d^T I d \leq 1\}.$$

However, the Euclidean ball is but one metric, and we might instead be more general, and choose to restrict d to lie in an ellipsoid

$$E_Q = \{d \in \Re^n \mid d^T Q d \leq 1\},$$

where Q is a given symmetric and *positive-definite* matrix. This leads to the more general direction-finding problem:

$$\text{minimize} \quad \nabla f(\bar{x})^T d$$

$$\text{s.t.} \quad Ad = 0$$
$$d^T Q d \leq 1.$$

The **projected steepest descent algorithm** is:

Step 1. $\bar{x}$ satisfies $A\bar{x} = b$. Compute $\nabla f(\bar{x})$.

Step 2. Solve the **direction-finding problem** (DFP):

DFP:
$$\bar{d} = \text{arg minimum} \quad \nabla f(\bar{x})^T d$$

$$\text{s.t.} \quad Ad = 0$$
$$d^T Q d \leq 1,$$

If $\nabla f(\bar{x})^T \bar{d} = 0$, stop. In this case, $\bar{x}$ is a Karush-Kuhn-Tucker point.

Step 3. Solve $\min_\alpha f(\bar{x} + \alpha \bar{d})$ for the stepsize $\bar{\alpha}$, perhaps chosen by an exact or inexact linesearch.

Step 4. Set $\bar{x} \leftarrow \bar{x} + \bar{\alpha} \bar{d}$. Go to **Step 1.**

Note that if $Q = I$ and the equality constraints $Ax = b$ are absent, this algorithm is just the steepest descent algorithm.

3 Properties of the Projected Steepest Descent Direction $\bar{d}$

Note that DFP is a convex program, and $d = 0$ is a Slater point. Therefore, the Karush-Kuhn-Tucker conditions are necessary and sufficient for optimality in DFP. These conditions are:

$$A\bar{d} = 0 \tag{1}$$

$$\bar{d}^T Q \bar{d} \leq 1 \tag{2}$$

$$\nabla f(\bar{x}) + A^T \bar{\pi} + 2\bar{\beta} Q \bar{d} = 0 \tag{3}$$

$$\bar{\beta} \geq 0 \tag{4}$$

$$\bar{\beta}(1 - \bar{d}Q\bar{d}) = 0. \tag{5}$$

As it turns out, it is extremely easy to solve these equations. (We will see this shortly.) Let $\bar{d}$ solve the equations (1)-(5) with multipliers $\bar{\beta}$ and $\bar{\pi}$.

Proposition 3.1 *$\bar{x}$ is a Karush-Kuhn-Tucker point of P if and only if $\nabla f(\bar{x})^T \bar{d} = 0$.*

Proposition 3.2 *$\bar{x}$ is a Karush-Kuhn-Tucker point of P if and only if $\bar{\beta} = 0$.*

Proposition 3.3 *If $\bar{x}$ is not a Karush-Kuhn-Tucker point of P, then $\bar{d}$ is a descent direction.*

Proposition 3.4 *The projected steepest descent algorithm has the same convergence properties and the same linear convergence as the steepest descent algorithm. Under the same conditions as in the steepest descent algorithm, the iterates converge to a Karush-Kuhn-Tucker point of P, and the convergence rate is linear, with a convergence constant that is bounded in terms of eigenvalues identically as in the steepest descent algorithm.*

4   Solving the Direction-Finding Problem (DFP)

Approach 1 to solving DFP: solving linear equations

Create the system of linear equations:

$$Q\tilde{d} + A^T\tilde{\pi} = -\nabla f(\bar{x}) \tag{6}$$

$$A\tilde{d} = 0 \tag{7}$$

and solve this linear system for $(\tilde{d}, \tilde{\pi})$ by any method at your disposal.

If $Q\tilde{d} = 0$, then

$$\nabla f(\bar{x}) + A^T \tilde{\pi} = 0$$

and so $\bar{x}$ is a Karush-Kuhn-Tucker point of P.

If $Q\tilde{d} \neq 0$, then rescale the solution as follows:

$$\bar{d} = \frac{\tilde{d}}{\sqrt{\tilde{d}^T Q \tilde{d}}},$$

$$\bar{\pi} = \tilde{\pi},$$

$$\bar{\beta} = \frac{1}{2\sqrt{\tilde{d}^T Q \tilde{d}}}.$$

Proposition 4.1 $(\bar{d}, \bar{\pi}, \bar{\beta})$ *defined above satisfy* (1), (2), (3), (4), *and* (5).

8

Note that the rescaling step is not necessary in practice, since we use a line-search.

Approach 2 to solving DFP: Formulas

Let

$$P = [Q^{-1} - Q^{-1}A^T(AQ^{-1}A^T)^{-1}AQ^{-1}]$$

$$\bar{\beta} = \frac{\sqrt{(\nabla f(\bar{x}))^T P(\nabla f(\bar{x}))}}{2}$$

$$\bar{\pi} = -(AQ^{-1}A^T)^{-1}AQ^{-1}(\nabla f(\bar{x}))$$

If $\bar{\beta} > 0$, let

$$\bar{d} = \frac{-P(\nabla f(\bar{x}))}{\sqrt{(\nabla f(\bar{x}))^T P(\nabla f(\bar{x}))}}.$$

If $\bar{\beta} = 0$, let $\bar{d} = 0$.

Proposition 4.2 *P is symmetric and positive semi-definite. Hence $\bar{\beta} \geq 0$.*

Proposition 4.3 *$(\bar{d}, \bar{\pi}, \bar{\beta})$ defined above satisfy (1), (2), (3), (4), and (5).*

9

5 Modification of Newton's Method with Linear Equality Constraints

Here we consider the following problem:

$$(\text{P:}) \qquad \text{minimize}_x \qquad f(x)$$

$$\text{s.t.} \qquad Ax = b.$$

Just as in the regular version of Newton's method, we approximate the objective with the quadratic expansion of $f(x)$ at $x = \bar{x}$:

$$(\tilde{\text{P}}:) \qquad \text{minimize}_x \quad h(x) := f(\bar{x}) + \nabla f(\bar{x})^T(x - \bar{x}) + \tfrac{1}{2}(x - \bar{x})^t H(\bar{x})(x - \bar{x})$$

$$\text{s.t.} \qquad\qquad\qquad Ax = b.$$

Now we solve this problem by applying the KKT conditions, and so we solve the following system for (x, u):

$$
\begin{aligned}
Ax && = \ b \\
\nabla h(x) \ + A^T u \ &= \ 0 \ .
\end{aligned}
$$

Now let us substitute the fact that:

$$\nabla h(x) = \nabla f(\bar{x}) + H(\bar{x})(x - \bar{x}) \quad \text{and} \quad A\bar{x} = b.$$

Substituting this and replacing $d = x - \bar{x}$, we have the system:

$$
\begin{aligned}
Ad & = 0 \\
H(\bar{x})d + A^T u & = -\nabla f(\bar{x}) \ .
\end{aligned}
$$

The solution (d, u) to this system yields the Newton direction d at $\bar{x}$.

Notice that there is actually a closed form solution to this system, if we want to pursue this route. It is:

$$
\begin{aligned}
d & = -H(\bar{x})^{-1}\nabla f(\bar{x}) + H(\bar{x})^{-1}A^T \left(AH(\bar{x})^{-1}A^T \right)^{-1} AH(\bar{x})^{-1}\nabla f(\bar{x}) \\
u & = -\left(AH(\bar{x})^{-1}A^T \right)^{-1} AH(\bar{x})^{-1}\nabla f(\bar{x}) \ .
\end{aligned}
$$

This leads to the following version of Newton's method for linearly constrained problems:

Newton's Method for Linearly Constrained Problems:

Step 0   Given x^0 for which $Ax^0 = b$, set $k \leftarrow 0$

Step 1 $\bar{x} \leftarrow x^k$. Solve for $(\bar{d}, \bar{u})$:

$$
\begin{aligned}
A\bar{d} & = 0 \\
H(\bar{x})\bar{d} + A^T\bar{u} & = -\nabla f(\bar{x}) \ .
\end{aligned}
\tag{8}
$$

If $H(\bar{x})\bar{d} = 0$, then stop.

Step 2   Choose step-size $\alpha^k = 1$.

Step 3   Set $x^{k+1} \leftarrow \bar{x} + \alpha^k\bar{d}$, $k \leftarrow k+1$. Go to **Step 1**.

Note the following:

- If $H(\bar{x})\bar{d} = 0$, then $\bar{x}$ is a KKT point. To see this, note from **Step 1** that $\nabla f(\bar{x}) + A^T \bar{u} = 0$, which are precisely the KKT conditions for this problem.

- Equations (8) are called the "normal equations". They were derived presuming that $H(\bar{x})$ is positive-definite, but can be used even when $H(\bar{x})$ is not positive-definite.

- If $H(\bar{x})$ is positive definite, then $\bar{d}$ is a descent direction. To see this, note that $\nabla f(\bar{x})^T \bar{d} = -\bar{d}^T H(\bar{x})\bar{d} < 0$ from (8) since $H(\bar{x})$ is positive definite.

6   The Variable Metric Method

In the projected steepest descent algorithm, the direction d must lie in the ellipsoid

$$E_Q = \{d \in \Re^n \mid d^T Q d \leq 1\},$$

where Q is fixed for all iterations. In a variable metric method, Q can vary at every iteration. The **variable metric algorithm** is:

Step 1. $\bar{x}$ satisfies $A\bar{x} = b$. Compute $\nabla f(\bar{x})$.

Step 2. Choose a postive-definite symmetric matrix Q. (Perhaps $Q = Q(\bar{x})$, i.e., the choice of Q may depend on the current point.) Solve the **direction-finding problem** (DFP):

DFP:
$$\bar{d} = \text{arg minimum} \quad \nabla f(\bar{x})^T d$$

$$\text{s.t.} \quad Ad = 0$$
$$d^T Q d \le 1,$$

If $\nabla f(\bar{x})^T \bar{d} = 0$, stop. In this case, $\bar{x}$ is a KKT point.

Step 3. Solve $\min_\alpha f(\bar{x} + \alpha \bar{d})$ for the stepsize $\bar{\alpha}$, perhaps chosen by an exact or inexact linesearch.

Step 4. Set $\bar{x} \leftarrow \bar{x} + \bar{\alpha}\bar{d}$. Go to **Step 1.**

All properties of the projected steepest descent algorithm still hold here.

Some strategies for choosing Q at each iteration are:

- $Q = I$

- Q is a given matrix held contstant over all iterations

- $Q = H(\bar{x})$ where $H(x)$ is the Hessian of $f(x)$. It is easy to show that that in this case, the variable metric algorithm is equivalent to Newton's method with a line-search, see the proposition below.

- $Q = H(\bar{x}) + \delta I$, where δ is chosent to be large for early iterations, but δ is chosen to be small for later iterations. One can think of this strategy as approximating the projected steepest descent algorithm in early iterations, followed by approximating Newton's method in later iterations.

Proposition 6.1 *Suppose that $Q = H(\bar{x})$ in the variable metric algorithm. Then the direction $\bar{d}$ in the variable metric method is a positive scalar times the Newton direction.*

  Proof: If $Q = H(\bar{x})$, then the vector $\bar{d}$ of the variable metric method is the optimal solution of DFP:

DFP:
$$\bar{d} = \arg \text{ minimum} \qquad \nabla f(\bar{x})^T d$$

$$\text{s.t.} \qquad\qquad Ad = 0$$
$$d^T H(\bar{x})d \le 1.$$

The Newton direction $\tilde{d}$ for P at the point $\bar{x}$ is the solution of the following problem:

NDP:
$$\hat{d} = \arg \text{ minimum} \quad \nabla f(\bar{x})^T d + \tfrac{1}{2}d^T H(\bar{x})d$$

$$\text{s.t.} \qquad\qquad Ad = 0.$$

If you write down the Karush-Kuhn-Tucker conditions for each of these two problems, you then can easily verify that $\bar{d} = \gamma\hat{d}$ for some scalar $\gamma > 0$.

q.e.d.