

NONLINEAR OPTIMIZATION WITH CONVEX CONSTRAINTS

The Goldstein-Levitin-Polyak algorithm

We consider an algorithm for solving the optimization problem under convex constraints. Although the convexity of the constraints is treated in its generality, in practice, convex constraints imply linear inequalities, especially upper and lower bounds on the decision variables. Apart from convexity, the main assumption on the constraints is that they are once differentiable.

The method belongs to the powerful class of algorithms developed by Goldstein-Levitin-Polyak. It requires that the constraints are satisfied at every iteration. This is an inconvenience in general. Yet, for linear constraints and for special constraint structures, it is simple to maintain feasibility. Two important advantage that seems to be offered in return for satisfying the constraints are the strong results concerning unit stepsize achievement and superlinear convergence. With regard to the latter, it is established that the *necessary and sufficient* condition for Q-superlinear rate of convergence is the two sided projected Hessian condition which, in other algorithms, can only ensure lesser Q-superlinear rates.

1. THE PROBLEM

Consider

$$\min \left\{ \mathcal{F}(x) \mid h(x) \leq 0 \right\}, \quad (1.1)$$

$\mathcal{F}: \mathbb{R}^n \rightarrow \mathbb{R}^1$, the elements of $h: \mathbb{R}^n \rightarrow \mathbb{R}^i$ are differentiable functions. The feasible region

$$\mathcal{R} = \left\{ x \in \mathbb{R}^n \mid h(x) \leq 0 \right\} \quad (1.2)$$

is assumed to be convex and $\mathcal{F}(x)$ is assumed to be bounded below on \mathcal{R} .

The convexity of the constraints allows the development of a special class algorithm. In particular, we use the fact that if two points belong to the convex set described by the constraints, then any point on the line segment joining the two points is also in the convex set. The point chosen by the algorithm on the line segment is the point that corresponds to an improved value of the objective function. The method we describe for solving (1.1) under this convexity assumption is a generalisation of the Goldstein-Levitin-Polyak (GLP) algorithm (Goldstein, 1964; Levitin and Polyak, 1966).

The basic GLP algorithm consists of the iterative scheme

$$x_{k+1} = P_{\mathcal{R}} \left(x_k + \tau_k d_k \right)$$

where d_k is a descent direction such as the steepest descent direction, $-\nabla \mathcal{F}(x_k)$. τ_k is the stepsize and $P_{\mathcal{R}}$ is the projection of $\left(x_k + \tau_k d_k \right)$ onto the feasible region \mathcal{R} .

2. THE ALGORITHM

Let the quadratic approximation be defined by

$$q_k(x) = \langle \nabla \mathcal{F}(x_k), x - x_k \rangle + \frac{1}{2} \langle x - x_k, \hat{Q}_k (x - x_k) \rangle \quad (2.1)$$

where \hat{Q}_k is a symmetric positive definite approximation to the Hessian Q_k of $\mathcal{F}(x)$ at x_k . The assumption that \hat{Q}_k is positive definite is not restrictive since it does not impair the convergence properties of the algorithm even when the true Hessian is not strictly positive definite. It will be shown in Section 4 that the convergence of the stepsizes, discussed below, to unity and that the Q-superlinear convergence rate depends on the accuracy of the projection of \hat{Q}_k onto the tangent manifold of the constraints.

The algorithm is based on the projections of the unconstrained step

$$\bar{x} = x_k - \alpha_k \hat{Q}_k^{-1} \nabla \mathcal{F}(x_k), \quad k = 0, 1, 2, \dots \quad (2.2)$$

with, $\alpha_k \in [1, \bar{\alpha}]$, $\bar{\alpha} \in [1, 2)$, onto the feasible region R using subproblem

$$\min \left\{ \frac{1}{2} \|x - \bar{x}\|_{\hat{Q}_k}^2 \mid x \in \mathcal{R} \right\} \quad (2.3, a)$$

where $\|v\|_{\hat{Q}}^2 = \langle v, \hat{Q} v \rangle$. **We note that for the choice $\alpha_k = 1$, reduces to the solution of**

$$\min \left\{ q_k(x) \mid x \in \mathcal{R} \right\}. \quad (2.3, b)$$

The lectures are mainly concerned with the choice $\alpha_k = 1$ and hence (2.3, b) but the results below are stated for the general case for $\alpha_k \in [1, \bar{\alpha}]$.

If x_k^p is the value of x that solves (2.3, a or b), then x_{k+1} is computed using

$$x_{k+1} = x_k + \tau_k (x_k^p - x_k) \quad (2.4)$$

with $\tau_k \in [0, 1]$ given by the smallest value of $j = 0, 1, 2, 3, \dots$, $\tau_k = (\pi)^j$, $\pi \in (0, 1)$, satisfying

$$\mathcal{F}(x_{k+1}) - \mathcal{F}(x_k) \leq \tau_k \rho_1 q_k(x_k^p), \quad \rho_1 \in (0, 1). \quad (2.5)$$

with ρ_1 an arbitrary number in the above range. An alternative to (2.5) is the Armijo-type stepsize strategy

$$\mathcal{F}(x_{k+1}) - \mathcal{F}(x_k) \leq \tau_k \rho_2 \langle \nabla \mathcal{F}(x_k), x_k^p - x_k \rangle, \quad \rho_2 \in \left(0, 1 - \frac{\bar{\alpha}}{2}\right). \quad (2.6)$$

with ρ_2 an arbitrary number in the above range.

The matrix \hat{Q}_k is approximated using Powell's (1978, b) modification to a **quasi-Newton** formula due to Broyden (1969; 1970), Fletcher (1970), Goldfarb (1970), Shanno

(1970) for approximating the Hessian of a Lagrangian. This modified BFGS (Broyden-Fletcher-Goldfarb-Shanno) formula is given by

$$\hat{Q}_{k+1} = \hat{Q}_k - \frac{\hat{Q}_k \delta_x \delta_x^T \hat{Q}_k}{\delta_x^T \hat{Q}_k \delta_x} + \frac{\nu \nu^T}{\delta_x^T \nu}. \quad (2.8,a)$$

where

$$\gamma_k = \nabla \mathcal{F}(x_{k+1}) - \nabla \mathcal{F}(x_k); \quad \delta_x = (x_{k+1} - x_k) \quad (2.8,b)$$

$$\nu = \theta \gamma_k + (1 - \theta) \hat{Q}_k \delta_x \quad (2.8,c)$$

$$\theta = \begin{cases} 1 & \text{if } \delta_x^T \gamma_k \geq 0.2 \delta_x^T \hat{Q}_k \delta_x \\ \frac{0.8 \delta_x^T \hat{Q}_k \delta_x}{\delta_x^T \hat{Q}_k \delta_x - \delta_x^T \gamma_k} & \text{if } \delta_x^T \gamma_k < 0.2 \delta_x^T \hat{Q}_k \delta_x \end{cases}. \quad (2.8,d)$$

For linearly constrained \mathcal{R} and positive definite \hat{Q}_k , (2.3) is a positive definite quadratic programming problem with a unique solution x_k^p . In the initial stages of the algorithm, $\alpha_k > 1$ provides the option for choosing large steps. As discussed in Sections 3 and 4 below, α_k does not affect the convergence properties of the algorithm, provided it is chosen the range $[1, \bar{\alpha}]$, with the sequence $\{\alpha_k\}$ converging to unity. An example is $\alpha_k = 2 - \tau_k$. It is shown below that $\tau_k > 0, \forall k$. Also, if $\tau_k = 1$ for $\alpha_k > 1$, then reducing α_{k+1} to unity makes it easier to maintain unit τ_k .

EXERCISE: Write in pseudocode the two versions of the GLP algorithm discussed in this section. Do a library search for suitable termination rules for the algorithm and incorporate them in the pseudocode.

3. CONVERGENCE

EXERCISE: Assuming \mathcal{R} is a system of linear inequalities, establish the descent property of the direction generated by the algorithm. (The descent property discussed below is for general convex feasible sets, which clearly include systems of inequalities. There is a direct way of establishing descent for linear inequalities.)

In this section the stepsize strategies (2.5) and (2.6) are justified for symmetric positive definite \hat{Q}_k and $\{\alpha_k\}$ satisfying the restrictions discussed in Section 2. It is shown that the algorithm is globally convergent. The main theorem of this section is (3.1) below which justifies (2.5)-(2.6) and establishes the existence of $\tau_k \in (0, 1]$ which satisfies the stepsize rules. Most subsequent results are dependent on Theorem 3.1.

Lemma 3.1

Let $\alpha_k \in (0, \infty)$, \hat{Q}_k be a bounded symmetric positive definite matrix and x_k be given by (2.2). Then, x_k^p solving (2.3) satisfies

$$\|x_k^p - x_k\|_{\hat{Q}_k}^2 \leq -\alpha_k \langle \nabla \mathcal{F}(x_k), x_k^p - x_k \rangle. \quad (3.1)$$

REMARK: If you have already done the exercise at the beginning of this section, you know this result holds for linear inequality constraints (for $\alpha_k = 1$, thence (2.3, b)). Hence, you can ignore the proof.

Proof

Since (2.3) is the projection of \bar{x} on the convex region \mathcal{R} , we have the inequality

$$\langle x_k^p - \mathcal{X}, \hat{Q}_k (x_k^p - \bar{x}) \rangle \leq 0 \quad (3.2)$$

for any $\mathcal{X} \in \mathcal{R}$ (see Rustem, 1998; Lemma 3.1.2). Thus, we have

$$\|x_k^p - x_k\|_{\hat{Q}_k}^2 \leq \langle x_k^p - x_k, \hat{Q}_k (x_k^p + \alpha_k \hat{Q}_k^{-1} \nabla \mathcal{F}(x_k) - x_k) \rangle - \alpha_k \langle \nabla \mathcal{F}(x_k), x_k^p - x_k \rangle.$$

The result follows as the first term on the right may be formulated as (3.2), for $\mathcal{X} = x_k$. \square

Lemma 3.2

Let $\tau_k \in [0, 1]$, \hat{Q}_k be symmetric positive definite and x_{k+1} be given by (2.4). Then

$$q_k(x_{k+1}) \leq \tau_k q_k(x_k^p) \quad (3.3)$$

Proof

$$q_k(x_{k+1}) = \tau_k \langle \nabla \mathcal{F}(x_k), x_k^p - x_k \rangle + \frac{1}{2} \tau_k^2 \langle x_k^p - x_k, \hat{Q}_k (x_k^p - x_k) \rangle \leq \tau_k q_k(x_k^p). \quad \square$$

Lemma 3.3

Let $\xi_{\bar{\alpha}} > 0$, $\alpha_k \in (0, \bar{\alpha}]$, $\bar{\alpha} \in (0, 2)$, \hat{Q}_k be bounded symmetric positive definite and x_k be given by (2.2). Then, x_k^p solving (2.3) satisfies

$$q_k(x_k^p) \leq \left(\frac{1}{2} - \frac{1}{\alpha_k} \right) \|x_k^p - x_k\|_{\hat{Q}_k}^2 \leq 0 \quad (3.4)$$

$$\xi_{\bar{\alpha}} \|x_k^p - x_k\|_{\hat{Q}_k}^2 \leq -q_k(x_k^p). \quad (3.5)$$

Proof

Lemma 3.1, yields (3.4) immediately. Also, (3.5) holds for $\xi_{\bar{\alpha}} = \left(\frac{1}{\bar{\alpha}} - \frac{1}{2} \right)$. \square

Theorem 3.1

Let

- (i) $\mathcal{R} \subseteq \mathbb{R}^n$ be a convex set and $\mathcal{F}(x) \in \mathbb{C}^2$,
- (ii) \hat{Q}_k in (2.1)-(2.3) satisfy
- m $\|\mathcal{X}\|^2 \leq \langle \mathcal{X}, \hat{Q}_k \mathcal{X} \rangle \leq M \|\mathcal{X}\|^2$, $\forall k$, $0 \neq \mathcal{X} \in \mathbb{R}^n$, $\infty > M \geq m > 0$,
- (iii) $\{\alpha_k\}$ be any sequence converging to unity, with $\alpha_k \in [0, \bar{\alpha}]$, $\bar{\alpha} \in [1, 2)$.

Then there exists a $\tau_k \in (0, 1]$ that satisfies (2.5) or (2.6) and hence the sequence $\{x_k\}$ computed by (2.4) generates a monotonically decreasing sequence $\{\mathcal{F}(x_k)\}$.

IF YOU UNDERSTAND THE DESCENT PROPERTY OF LEMMA 3.1 THE PROOF IS EASY TO ESTABLISH.

Remark.

Second order Taylor series expansion of a function $f: \mathbb{R}^n \rightarrow \mathbb{R}^l$, and $f \in \mathbb{C}^2$, for any $x, d \in \mathbb{R}^n, \theta \in \mathbb{R}^l$, is given by

$$f(x + \theta d) = f(x) + \theta \langle \nabla f(x), d \rangle + \theta^2 \left(\int_0^1 (1-t) \langle d, \nabla^2 f(x + t\theta d) \rangle dt \right).$$

Proof

Using the second order Taylor series expansion,

$$\begin{aligned} \mathcal{F}(x_{k+1}) - \mathcal{F}(x_k) &= \langle \nabla \mathcal{F}(x_k), x_{k+1} - x_k \rangle + \frac{1}{2} \langle x_{k+1} - x_k, \hat{Q}_k (x_{k+1} - x_k) \rangle \\ &+ \int_0^1 (1-t) \langle x_{k+1} - x_k, \left\{ \mathcal{Q}(x_k(t)) - \hat{Q}_k \right\} (x_{k+1} - x_k) \rangle dt \end{aligned} \tag{3.7}$$

$$\leq \tau_k q_k(x_k^p) + \tau_k^2 \beta_k \|x_k^p - x_k\|^2 \tag{3.8}$$

where $x_k(t) = x_k - t(x_k - x_{k+1})$, $\mathcal{Q}(\cdot)$ is the Hessian of \mathcal{F} at (\cdot) ,

$$\beta_k = \int_0^1 (1-t) \|\mathcal{Q}(x_k(t)) - \hat{Q}_k\| dt,$$

and (3.3) is applied to obtain (3.8). From (3.5) and (3.8), we have

$$\mathcal{F}(x_{k+1}) - \mathcal{F}(x_k) \leq \tau_k q_k(x_k^p) \left(1 - \frac{\tau_k \beta_k}{m \xi_\alpha} \right) \tag{3.9}$$

Since $0 < \rho_1 < 1$, there is a $\tau_k \in (0, 1]$ such that

$$\rho_1 < 1 - \frac{\tau_k \beta_k}{m \xi_\alpha} \leq 1. \tag{3.10}$$

By Lemma (3.3), $q_k(x_k^p) \leq 0$. Thus, (2.5) holds for this τ_k . Suppose τ^0 is the largest $\tau \in [0, 1]$ satisfying inequality (2.5). All $\tau \leq \tau^0$ also satisfy this condition and the selected $\tau_k \in [\tau \tau^0, \tau^0]$. It follows that $\{\mathcal{F}(x_k)\}$ is monotonically decreasing.

To show that $\{\mathcal{F}(x_k)\}$ is monotonically decreasing for (2.6), we use (3.1) with (3.7) to yield

$$\mathcal{F}(x_{k+1}) - \mathcal{F}(x_k) \leq \tau_k \langle \nabla \mathcal{F}(x_k), x_k^p - x_k \rangle \left(1 - \frac{\tau_k \alpha_k}{2} - \frac{\tau_k \beta_k \alpha_k}{m} \right)$$

Since $0 < \rho_2 \leq 1 - \frac{\bar{\alpha}}{2}$, there is a $\tau_k \in (0, 1]$ such that (3.11)

$$0 < \rho_2 \leq 1 - \frac{\tau_k \alpha_k}{2} - \frac{\tau_k \beta_k \alpha_k}{m} \leq 1 - \frac{\bar{\alpha}}{2}. \quad (3.12)$$

By Lemma (3.1), $\langle \nabla \mathcal{F}(x_k), x_k^p - x_k \rangle \leq 0$. Thus, (2.6) holds for this τ_k . Suppose τ^0 is the largest $\tau \in [0, 1]$ satisfying inequality (2.6). All $\tau \leq \tau^0$ also satisfy this condition and the selected $\tau_k \in [\pi\tau^0, \tau^0]$. It follows that $\{\mathcal{F}(x_k)\}$ is monotonically decreasing. \square

THEOREM 3.2 BELOW IS AN EXAMPLE FOR THE USE OF THE BOLZANO-WEIERSTRASS THEOREM. YOU ARE NOT RESPONSIBLE FOR ANY PROOFS FROM THIS POINT ON IN THIS SET OF NOTES.

Theorem 3.2

Let

- (i) the assumptions of Theorem 3.1 be satisfied,
- (ii) the set $\mathbb{L} = \left\{ x \in \mathcal{R} \mid \mathcal{F}(x) \leq \mathcal{F}(x_0) \right\}$ be bounded.

Then, we have

$$\lim_{k \rightarrow \infty} q_k(x_k^p) = 0 \quad (3.13)$$

and

$$\lim_{k \rightarrow \infty} \langle \nabla \mathcal{F}(x_k), x_k^p - x_k \rangle = 0. \quad (3.14)$$

Proof

Given $0 < \rho_1 < 1$, by (3.9), the choice $\tau^0 = \min \left\{ 1, \frac{m \xi_{\bar{\alpha}}(1 - \rho_1)}{\beta_k} \right\}$ always satisfies the stepsize strategy (2.5). Clearly, τ_k , chosen as $\tau_k = \pi^{jk}$, as discussed above, is in the range $\tau_k \in [\pi\tau^0, \tau^0]$ and thereby also satisfies (2.5). As $\mathcal{F}(x) \in \mathbb{C}^2$, and \mathbb{L} is compact, there is a scalar $\bar{M} < \infty$ such that $\beta_k \leq \bar{M}$. Similarly, given $0 < \rho_2 < 1 - \frac{\bar{\alpha}}{2}$, (2.6) is satisfied by $\tau^0 = \min \left\{ 1, (1 - \rho_2) \left(\bar{\alpha} \left(\frac{1}{2} + \frac{\beta_k}{m} \right)^{-1} \right) \right\}$. As $m, \bar{\alpha}, \xi_{\bar{\alpha}} \geq 0$, we have established that there is an $\epsilon > 0$ such that the stepsizes determined by (2.5) or (2.6) satisfy $\tau_k \geq \epsilon > 0, \forall k$.

The boundedness of \mathcal{F} on \mathbb{L} and $q_k(x_k^p) \leq 0$, imply, in the case of (2.5), that

$$0 \leq \rho_1 \sum_k \tau_k |q_k(x_k^p)| \leq \sum_k \mathcal{F}(x_k) - \mathcal{F}(x_{k+1}) < \infty.$$

Since $\tau_k \geq \epsilon > 0$, this yields (3.13). (3.14) is established similarly by using (2.6) and $\langle \nabla \mathcal{F}(x_k), x_k^p - x_k \rangle \leq 0$. \square

Lemma 3.4

If (3.13) or if (3.14) are satisfied, then

$$\lim_{k \rightarrow \infty} \|x_k^p - x_k\| = 0. \quad (3.15)$$

Proof

The result follows from Lemmas (3.3) and (3.4) for (3.13) and (3.14) respectively. \square

Theorem 3.3

Let

- (i) the assumptions of Theorem 3.1 be satisfied,
- (ii) the assumptions of Theorem 3.2 be satisfied,
- (iii) $h(x)$ be once differentiable,
- (iv) the active constraint gradients at x_k^p are linearly independent (alternatively, instead of the linear independence condition, it can be assumed that the multipliers associated with (2.3) at x_k^p are bounded (Fiacco and McCormick, 1968))

Then,

- (a) the algorithm in Section 2, with stepsize strategy (2.5) or (2.6), generates a sequence $\{x_k\}$ that converges to x_* , and
- (b) if, furthermore, strict complementarity holds at the solution of subproblem (2.3), for large k , μ_k predicts the active inequality constraints at x_* .

Proof

By Theorem 3.1, the algorithm ensures the decrease of $\mathcal{F}(x)$ at each iteration, thereby ensuring $x_k \in \mathbb{F}$, for \mathbb{F} compact. Hence, $\{x_k\}$ has a limit point, x_* . Without loss of generality, we can take $\{x_k\} \rightarrow x_*$. To show that x_* also satisfies the necessary conditions for optimality for problem (1.1), we consider the optimality of subproblem (2.3). Let μ_{k+1} and ∇h_k^p denote respectively the multipliers of the subproblem and the matrix whose columns are the constraint gradients at x_k^p . Using (2.2) and (2.3), the conditions for (2.3) are given by

$$\nabla \mathcal{F}(x_k) + \frac{1}{\alpha_k} \hat{Q}_k (x_k^p - x_k) + \nabla h_k^p \mu_{k+1} = 0 \quad (3.16, a)$$

$$h(x_k^p) \leq 0; \quad \langle h(x_k^p), \mu_{k+1} \rangle = 0; \quad \mu_{k+1} \geq 0. \quad (3.16, b)$$

Using (3.15), we have $\lim_{k \rightarrow \infty} x_k = \lim_{k \rightarrow \infty} x_k^p = x_*$ and $\nabla \mathcal{F}(x_*) + \nabla h_* \mu_* = 0$.

To show (b) we note that, in view of the last two conditions in (3.16, b), for sufficiently large k and with strict complementarity holding, none of the inactive constraints, i.e.

$$h_k^j < 0; \quad \mu_{k+1}^j = 0 \quad (3.10)$$

are predicted to be active at x_* , $h_*^j(x_*) < 0$, $\mu_*^j = 0$. □

4. UNIT STEPSIZES AND SUPERLINEAR CONVERGENCE RATES

We consider the convergence of τ_k to unity. Coupled with $\alpha_k \rightarrow 1$, this leads to the Q-superlinear convergence rate of the algorithm. It is shown that both the convergence τ_k to 1 and superlinear convergence are dependent on the Hessian approximation \hat{Q}_k . The algorithm uses strictly positive definite \hat{Q}_k even when the Hessian at the solution is not strictly positive definite. First, we establish the subspace in which \hat{Q}_k needs to approximate the Hessian of $\mathcal{F}(x)$. This is used in Theorem 4.1 to establish the attainment of unit stepsizes and in Theorem 4.2 to establish the convergence rate.

A consequence of Theorem 3.3 is that, for sufficiently large k , with strict complementarity holding, μ_k predicts the constraints active at x^* . Thus, for large k , the constraints satisfied as strict inequalities at x^* do not affect the computation of x_k . At that stage, it would make no difference if the constraints satisfied as equalities at x^* were treated as equality constraints. Hence, for large k , $h^j(x^*) = 0$ implies $h^j(x_k) = 0$ and $h^j(x^*) < 0$ implies $h^j(x_k) < 0$. Let h^* denote the vector constraints satisfied as equalities at x^* . Thus, we have $h^*(x_k) = 0$, for large k .

In order to characterize the subspace for approximating the Hessian of $\mathcal{F}(x)$, we need to invoke a special case of the mean-value theorem that holds for h^* (Ortega and Rheinboldt, 1970).

Remark: Mean-Value Theorems

We note that mean value theorems usually apply to mappings $f: \mathbb{F} \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ and do not in general hold for mappings $f: \mathbb{F} \subset \mathbb{R}^n \rightarrow \mathbb{R}^m, m > 1$. We construct the result by treating each element of the vector h^* individually. \square

Provided each element of the vector h^* is differentiable, on an open convex set $\mathbb{H}_0 \subset \mathbb{H}$, then for any two points $x_{k+1}, x_k \in \mathbb{H}_0$, there exist $t_1, t_2, \dots, t_j, \dots \in (0, 1)$ such that

$$\mathcal{H}(x_{k+1}, x_k) = \left[\nabla h^{*1}(x_k + t_1(x_{k+1} - x_k)), \nabla h^{*2}(x_k + t_1(x_{k+1} - x_k)), \dots \right] \tag{4.1}$$

where ∇h^{*j} denotes the gradient of the j^{th} element of h^* , and

$$h^*(x_{k+1}) - h^*(x_k) = \mathcal{H}^T(x_{k+1}, x_k) (x_{k+1} - x_k). \tag{4.2}$$

Let the matrix \mathcal{H}_k have the same rank as $\mathcal{H}(x_{k+1}, x_k)$ and let a linearly independent subset of the columns of $\mathcal{H}(x_{k+1}, x_k)$ form the columns of \mathcal{H}_k . Thus \mathcal{H}_k is of full rank and we can form the operator $P_k = I - \mathcal{H}_k(\mathcal{H}_k^T \mathcal{H}_k)^{-1} \mathcal{H}_k^T$, which satisfies $P_k P_k = P_k$ and

$$P_k (x_{k+1} - x_k) = x_{k+1} - x_k. \tag{4.3}$$

As $\{x_k\} \rightarrow x^*$,

$$\left\{ \mathcal{H}(x_{k+1}, x_k) \right\} \rightarrow \left[\nabla h^{*1}(x^*), \nabla h^{*2}(x^*), \dots \right] \tag{4.4}$$

and P_k becomes the operator projecting vectors in \mathbb{R}^n onto the active constraints at x^* .

Theorem 4.1

Let the assumptions of Theorem 3.3 be satisfied. Then, there is a $\sigma > 0$, such that

(a)

$$\frac{\|(\mathcal{Q}^* - \hat{\mathcal{Q}}_k)(x_{k+1} - x_k)\|}{\|x_{k+1} - x_k\|} \leq \sigma, \tag{4.5}$$

or,

(b) for large k ,

$$\frac{\|P_k(\mathcal{Q}^* - \hat{\mathcal{Q}}_k)P_k(x_{k+1} - x_k)\|}{\|x_{k+1} - x_k\|} \leq \sigma, \tag{4.6}$$

we have $\{\tau_k\} \rightarrow 1$.

Remark. Inequality (4.5) requires \hat{Q}_k to be close to Q_* . This may be satisfied for Q_* and \hat{Q}_k strictly positive definite or for some \hat{Q}_k positive definite and Q_* not positive definite. If Q_* is not positive definite, and there is no \hat{Q}_k satisfying (4.5), we need to consider (4.6). The latter demands only the projections of Q_* , \hat{Q}_k , on the constraints, to be close and the projection of Q_* to be positive definite.

Proof

For stepsize (2.5), we can write (3.7) as

$$\begin{aligned} \mathcal{F}(x_{k+1}) - \mathcal{F}(x_k) &\leq \tau_k q_k(x_k^p) \\ &+ \int_0^1 (1-t) \langle x_{k+1} - x_k, \{ Q(x_k(t)) - Q_* + Q_* - \hat{Q}_k \} (x_{k+1} - x_k) \rangle dt \end{aligned} \quad (4.7)$$

$$\begin{aligned} &\leq \tau_k q_k(x_k^p) + \tau_k^2 \beta_k^* \|x_k^p - x_k\|^2 \\ &\quad + \frac{\tau_k^2}{2} \|x_k^p - x_k\| \left\| \left(Q_* - \hat{Q}_k \right) \left(x_k^p - x_k \right) \right\| \end{aligned} \quad (4.8)$$

$$\leq \tau_k q_k(x_k^p) \left[1 - \frac{\tau_k}{m\xi_\alpha} \left[\beta_k^* + \frac{1}{2} \frac{\| (Q_* - \hat{Q}_k) (x_{k+1} - x_k) \|}{\|x_{k+1} - x_k\|} \right] \right] \quad (4.9)$$

where $\beta_k^* = \int_0^1 (1-t) \| Q(x_k(t)) - Q_* \| dt$ and (4.9) is obtained by invoking Lemma (3.3). The scalar $\rho_1 \in (0, 1)$ in (2.5) requires τ_k to satisfy

$$\rho_1 \leq 1 - \frac{\tau_k}{m\xi_\alpha} \left[\beta_k^* + \frac{1}{2} \frac{\| (Q_* - \hat{Q}_k) (x_{k+1} - x_k) \|}{\|x_{k+1} - x_k\|} \right] \leq 1. \quad (4.10)$$

As in (3.9)-(3.10), there exists a $\tau_k \in (0, 1]$, satisfying (4.9) and hence (2.5). If σ in (4.5) is such that

$$\frac{1}{m\xi_\alpha} \left[\beta_k^* + \frac{1}{2} \sigma \right] \leq 1 - \rho_1 \quad (4.11)$$

(in view of $\{x_k\} \rightarrow x_*$, $\beta_k^* \rightarrow 0$, this defines the number σ) then (4.10) holds with $\tau_k=1$, and therefore, because $q_k(x_k^p) \leq 0$, (2.5) is satisfied with $\tau_k = 1$.

If (4.5) cannot be achieved because Q_* is not positive definite, then the projection of $Q_* - \hat{Q}_k$ can be used, for large k , by invoking (4.4) in (4.8) to yield

$$\mathcal{F}(x_{k+1}) - \mathcal{F}(x_k) \leq \tau_k q_k(x_k^p) \left[1 - \frac{\tau_k}{m\xi_\alpha} \left[\beta_k^* + \frac{1}{2} \frac{\| P_k(Q_* - \hat{Q}_k) P_k(x_{k+1} - x_k) \|}{\|x_{k+1} - x_k\|} \right] \right].$$

(4.12)

Using (4.6) and the same arguments as before, we establish that (2.5) is satisfied with $\tau_k = 1$.

For stepsize (2.6), (3.7) with Lemma (3.1) yields

$$\begin{aligned} \mathcal{F}(x_{k+1}) - \mathcal{F}(x_k) &\leq \tau_k \langle \nabla \mathcal{F}(x_k), x_k^p - x_k \rangle \\ &\times \left[1 - \frac{\bar{\alpha} \tau_k}{2} - \frac{\bar{\alpha} \tau_k}{m} \left[\beta_k^* + \frac{1}{2} \frac{\|(\mathcal{Q}^* - \hat{\mathcal{Q}}_k)(x_{k+1} - x_k)\|}{\|x_{k+1} - x_k\|} \right] \right] \end{aligned} \quad (4.13)$$

The scalar $\rho_2 \in \left(0, 1 - \frac{\bar{\alpha}}{2} \right)$ in (2.6) requires τ_k to satisfy

$$\rho_2 \leq 1 - \frac{\bar{\alpha} \tau_k}{2} - \frac{\bar{\alpha} \tau_k}{m} \left[\beta_k^* + \frac{1}{2} \frac{\|(\mathcal{Q}^* - \hat{\mathcal{Q}}_k)(x_{k+1} - x_k)\|}{\|x_{k+1} - x_k\|} \right] \leq 1 - \frac{\bar{\alpha}}{2}. \quad (4.14)$$

As in (3.11)-(3.12), there exists a $\tau_k \in (0, 1]$, satisfying (2.6). If σ in (4.5) is such that

$$\frac{\bar{\alpha}}{m} \left[\beta_k^* + \frac{1}{2} \sigma \right] \leq 1 - \frac{\bar{\alpha}}{2} - \rho_1 \quad (4.15)$$

(since $\{x_k\} \rightarrow x^*$, $\beta_k^* \rightarrow 0$, this defines σ) then (4.14) holds with $\tau_k=1$, and therefore, because $\langle \nabla \mathcal{F}(x_k), x_k^p - x_k \rangle \leq 0$, (2.6) is satisfied with $\tau_k = 1$.

If (4.5) cannot be achieved because \mathcal{Q}^* is not positive definite, then the projection of $\mathcal{Q}^* - \hat{\mathcal{Q}}_k$ can be used, for large k , by invoking (4.4) to yield

$$\begin{aligned} \mathcal{F}(x_{k+1}) - \mathcal{F}(x_k) &\leq \tau_k \langle \nabla \mathcal{F}(x_k), x_k^p - x_k \rangle \\ &\times \left[1 - \frac{\bar{\alpha} \tau_k}{2} - \frac{\bar{\alpha} \tau_k}{m} \left[\beta_k^* + \frac{1}{2} \frac{\|P_k(\mathcal{Q}^* - \hat{\mathcal{Q}}_k)P_k(x_{k+1} - x_k)\|}{\|x_{k+1} - x_k\|} \right] \right]. \end{aligned} \quad (4.16)$$

Using (4.6) and the same arguments as before, we establish that (2.6) is satisfied with $\tau_k = 1$. \square

Remark. The above proof illustrates that it is easier to attain $\tau_k = 1$ for smaller values of $\bar{\alpha}$. $\{\tau_k\}$ will accelerate towards unity as $\{\alpha_k\} \rightarrow 1$. Furthermore, if $\tau_k = 1$ while $\alpha_k > 1$, then reducing α_k to unity in subsequent iterations will increase $\xi_{\bar{\alpha}}$ to its largest value $\xi_{\bar{\alpha}} = \frac{1}{2}$. Consider, for example, the effect of doing this in (4.11) or (4.17). As $\xi_{\bar{\alpha}}$ gets larger, the bound $m\xi_{\bar{\alpha}}$ on these inequalities also gets larger. The same also applies to the bound on σ . Thus, conditions (4.5) and (4.6) become easier to satisfy and this makes it easier to maintain $\tau_k=1$.

The left sides of bounds (4.5) and (4.6) are expected to approach zero if superlinear convergence is to be achieved. Theorem 4.2 below establishes these superlinear convergence conditions. In chapter 8, this discussion is revisited and further generalised for Quasi-Newton algorithms, following the earlier results of Dennis and Moré (1977), Han (1976) and Powell (1978, a), among others.

Lemma 4.1

Let $\{x_k\} \rightarrow x^*$ and $\|x^* - x_{k+1}\| \leq \hat{e} \|x^* - x_k\|$, for some $\hat{e} \in [0, \infty)$. Then

$$\|x_{k+1} - x_k\| \leq e_0 \|x^* - x_k\|. \tag{4.17}$$

where $e_0 = 1 + \hat{e}$.

Proof

The proof is immediate from the triangle inequality

$$\|x_{k+1} - x^* + x^* - x_k\| \leq \|x_{k+1} - x^*\| + \|x^* - x_k\| \leq (1 + \hat{e}) \|x_k - x^*\|. \quad \square$$

Remark. The hypothesis

$$\|x^* - x_{k+1}\| \leq \hat{e} \|x^* - x_k\|$$

is always satisfied whenever (4.17) is invoked below.

Definition. Let the sequence $\{x_k\} \rightarrow x^*$. If

$$\lim_{k \rightarrow \infty} \frac{\|x^* - x_{k+1}\|}{\|x^* - x_k\|} = 0 \tag{4.18}$$

then $\{x_k\}$ is convergent at a *Q-superlinear rate*.

Lemma 4.2

Let $\{x_k\} \rightarrow x^*$. Then $\{x_k\}$ is Q-superlinearly convergent, i.e.

$$\|x^* - x_{k+1}\| \leq r_k^{(0)} \|x^* - x_k\| ; \quad \lim_{k \rightarrow \infty} r_k^{(0)} = 0,$$

iff $\|x_{k+1} - x_k\| \leq r_k \|x_k - x_{k-1}\|$ with $\lim_{k \rightarrow \infty} r_k = 0$.

Proof

We have

$$\begin{aligned} \|x^* - x_k\| &\leq \lim_{t \rightarrow \infty} \sum_{j=k}^{t-1} \|x_{j+1} - x_j\| \\ &\leq r_k \|x_k - x_{k-1}\| (1 + \omega + \omega^2 + \omega^3 + \dots) \\ &\leq \frac{r_k}{1-\omega} \left\{ \|x_k - x^*\| + \|x^* - x_{k-1}\| \right\} \end{aligned}$$

for some $\omega \in [0, 1)$ and k sufficiently large. As $\{r_k\} \rightarrow 0$, ω is chosen such that $1 > r_k + \omega$, $\forall k \geq K_0$. K_0 is an integer and is such that $r_k < 1$, $\forall k \geq K_0$. Rearranging the above expression, yields the required result.

Suppose that $\|x_* - x_k\| \leq r_k^{(0)} \|x_* - x_{k-1}\|$, $\lim_{k \rightarrow \infty} r_k^{(0)} = 0$. The desired result is obtained using Lemma (2.1) and $r_k^{(0)} < 1$

$$\|x_* - x_k\| \leq r_k^{(0)} \left\{ \|x_* - x_k\| + \|x_k - x_{k-1}\| \right\} \leq \left(\frac{r_k^{(0)}}{1 - r_k^{(0)}} \right) \|x_k - x_{k-1}\| . \square$$

Theorem 4.2

Let

- (i) the assumptions of Theorem 3.3 be satisfied,
- (ii) k be large, such that, by Theorem 4.1, $\alpha_k = \tau_k = 1$.

Then, the sequence $\{x_k\}$ generated by the GLP algorithm in Section 2 converges at a Q-superlinear rate iff

$$\lim_{k \rightarrow \infty} \frac{\|P_k (\hat{Q}_* - \hat{Q}_k) P_k (x_{k+1} - x_k)\|}{\|x_{k+1} - x_k\|} = 0 \quad (4.19)$$

Proof

The first order expansion of $\nabla \mathcal{F}(x)$ can be written using (4.4) as

$$\begin{aligned} \nabla \mathcal{F}(x_k) &= \nabla \mathcal{F}(x_{k-1}) + \hat{Q}_k (x_k - x_{k-1}) \\ &+ \int_0^1 \left\{ \mathcal{Q} (x_{k-1}(t)) - \hat{Q}_{k-1} \right\} P_{k-1} (x_k - x_{k-1}) > dt. \end{aligned} \quad (4.20)$$

The gradient of the objective function in (2.3) is given by the first two terms on the right in (4.20). Thus, for $x_{k+1} \in \mathcal{R}$, the inequality

$$\langle \nabla \mathcal{F}(x_{k-1}) + \hat{Q}_k (x_k - x_{k-1}), x_{k+1} - x_k \rangle = \langle q_{k-1}(x_k), x_{k+1} - x_k \rangle \geq 0 \quad (4.21)$$

follows from the optimality of x_k in (2.3) for

$$\min \left\{ \frac{1}{2} \left\| x - x_{k-1} + \hat{Q}_{k-1}^{-1} \nabla \mathcal{F}(x_{k-1}) \right\|_{\hat{Q}_{k-1}}^2 \mid x \in \mathcal{R} \right\}.$$

Also, the inequality

$$q_k(x_{k+1}) \geq \langle \nabla \mathcal{F}(x_k), x_{k+1} - x_k \rangle \quad (4.22)$$

follows from the positive definiteness of \hat{Q}_k . For large k , using (2.4) with $\alpha_k = \tau_k = 1$, (4.20)-(4.22) and Lemma 3.3, we have

$$\frac{m}{2} \|x_{k+1} - x_k\|^2 \leq -q_k(x_{k+1})$$

$$\begin{aligned}
 &\leq - \langle \nabla \mathcal{F}(\mathbf{x}_k), \mathbf{x}_{k+1} - \mathbf{x}_k \rangle \\
 &\leq - \langle \mathbf{q}_{k-1}(\mathbf{x}_k), \mathbf{x}_{k+1} - \mathbf{x}_k \rangle \\
 &\quad - \langle \mathbf{x}_{k+1} - \mathbf{x}_k, \mathbf{P}_k \{ \mathcal{Q}^* - \hat{\mathcal{Q}}_{k-1} \} \mathbf{P}_{k-1} (\mathbf{x}_k - \mathbf{x}_{k-1}) \rangle \\
 &\quad - \int_0^1 \langle \mathbf{x}_{k+1} - \mathbf{x}_k, \{ \mathcal{Q}(\mathbf{x}_{k-1}(t)) - \mathcal{Q}^* \} (\mathbf{x}_k - \mathbf{x}_{k-1}) \rangle dt \\
 &\leq \left[\bar{\beta}_k + \frac{\| \mathbf{P}_k \{ \mathcal{Q}^* - \hat{\mathcal{Q}}_{k-1} \} \mathbf{P}_{k-1} (\mathbf{x}_k - \mathbf{x}_{k-1}) \|}{\| \mathbf{x}_k - \mathbf{x}_{k-1} \|} \right] \\
 &\quad \times \| \mathbf{x}_k - \mathbf{x}_{k-1} \| \| \mathbf{x}_{k+1} - \mathbf{x}_k \| \tag{4.23}
 \end{aligned}$$

where $\bar{\beta}_k = \int_0^1 \| \mathcal{Q}(\mathbf{x}_{k-1}(t)) - \mathcal{Q}^* \| dt$ and, as $\{\mathbf{x}_k\} \rightarrow \mathbf{x}^*$, $\{\bar{\beta}_k\} \rightarrow 0$. Hence, (4.23) yields

$$\| \mathbf{x}_{k+1} - \mathbf{x}_k \| \leq r_k \| \mathbf{x}_k - \mathbf{x}_{k-1} \|$$

where

$$\begin{aligned}
 r_k = \quad &\frac{2}{m} \left\{ \bar{\beta}_k + \frac{\| \mathbf{P}_{k-1} \{ \mathcal{Q}^* - \hat{\mathcal{Q}}_{k-1} \} \mathbf{P}_{k-1} (\mathbf{x}_k - \mathbf{x}_{k-1}) \|}{\| \mathbf{x}_k - \mathbf{x}_{k-1} \|} \right. \\
 &\left. + \frac{\| (\mathbf{P}_k - \mathbf{P}_{k-1}) \{ \mathcal{Q}^* - \hat{\mathcal{Q}}_{k-1} \} \mathbf{P}_{k-1} (\mathbf{x}_k - \mathbf{x}_{k-1}) \|}{\| \mathbf{x}_k - \mathbf{x}_{k-1} \|} \right\}.
 \end{aligned}$$

If (4.19) is satisfied, $\lim_{k \rightarrow \infty} r_k = 0$ and Lemma 4.2 yields the desired result.

Suppose, conversely, that $\{\mathbf{x}_k\}$ converges Q-superlinearly and thence, by Lemma 4.2, $\| \mathbf{x}_{k+1} - \mathbf{x}_k \| \leq r_k \| \mathbf{x}_k - \mathbf{x}_{k-1} \|$ with $\lim_{k \rightarrow \infty} r_k = 0$. By (4.23), we have (4.19). \square

REFERENCES

- Allwright, J.C. (1980). A Feasible Direction Algorithm for Convex Optimization: Global Convergence Rates, *J Optim Theory Appl*, 1-18.
- Apostol, T.M. (1981). *Mathematical Analysis*, Second Edition, Addison Wesley, Reading, Massachusetts.
- Bertsekas, D.P. (1976). On the Goldstein-Levitin-Polyak Gradient Projection Method, *IEEE Trans on Automatic Control*, AC-21, 174-183.

- Bertsekas, D.P. (1982). Projected Newton Methods for Optimization Problems with Simple Constraints, *SIAM J Control Optim*, 20, 221-246.
- Broyden C.G. (1969). "A New Method for Solving Nonlinear Simultaneous Equations", *Computer J.*, 12, 95-100.
- Broyden C.G. (1970). "The Convergence of a Class of Double-Rank Minimisation Algorithms 2. The new Algorithm", *J. Inst. Maths. Applics.*, 6, 222-231.
- Demyanov, V.F. and A.M. Rubinov (1971). *Approximate Methods in Optimization Problems*, American Elsevier, New York.
- Dennis, J.E. and J.J. Moré (1977). Quasi-Newton Methods, Motivation and Theory, *SIAM Review*, 19, 46-89.
- Dunn, J.C. (1979). Rates of Convergence for Conditional Gradient Algorithms Near Singular and Nonsingular Extremals, *SIAM J Control Optim*, 17, 187-211.
- Dunn, J.C. (1980). Newton's Method and the Goldstein Step-length Rule for Constrained Minimization Problems, *SIAM J Control Optim*, 18, 659-674.
- Dunn, J.C. (1981). Global and Asymptotic Convergence Rate Estimates for a Class of Projected Gradient Processes, *SIAM J Control Optim*, 19, 368-400.
- Fiacco, A.V. and G.P. McCormick (1968). *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*, Wiley, New York.
- Fletcher, R. (1970). "A New Approach to Variable Metric Algorithms", *Computer J.*, 13, 317-322.
- Gafni, E. and D. Bertsekas (1984). "Two-Metric Projection Methods for Constrained Optimization", *SIAM J. Control and Optimization*, 22, 936-964.
- Goldfarb D. (1970). "A Family of Variable Metric Algorithms Derived by Variational Means", *Mathematics of Computation*, 24, 23-26.
- Goldstein, A.A. (1964). Convex Programming in Hilbert Space, *Bulletin AMS*, 70, 709-710.
- Han, S-P (1976). Superlinearly Convergent Variable Metric Algorithms for General Nonlinear Programming Problems, *Math Programming*, 11, 263-282.
- Levitin, E.S. and B.T Polyak (1966). Constrained Minimization Methods, *USSR Comp Math and Math Phys*, 6, 1-50.
- McCormick, G.P. and R.A Tapia (1972). The Gradient Projection Method Under Mild Differentiability Conditions, *SIAM J Control Optim*, 10, 93-98

- Ortega, J.M. and W.C. Rheinboldt (1970). *Iterative Solution of Non-linear Equations in Several Variables*, Academic Press, London and New York
- Polak, E. (1971). *Computational Methods in Optimization*, Academic Press, London and New York
- Powell, M.J.D. (1978, a). The Convergence of Variable Metric Methods for Nonlinearly Constrained Optimization Calculations, In: J.B. Rosen, O.L. Mangasarian, K. Ritter (eds). *Nonlinear Programming*, Academic Press, New York:27-63
- Powell, M.J.D. (1978, b). "A fast Algorithm for Nonlinearly Constrained Optimization Calculations", *Numerical Analysis Proceedings, Dundee 1977*, Edited by G.A. Watson, Springer-Verlag, Berlin.
- Psenichny, B.N. and Y.M. Danilin (1978). *Numerical Methods in Extremal Problems*. Mir Publishers, Moscow
- Rustem, B. (1984). A Class of Superlinearly Convergent Projection Algorithms with Relaxed Stepsizes, *Appl Math Optim*, 12, 29-43.
- Shanno, D.F. (1970). "Conditioning of Quasi-Newton Methods for Function Minimization", *Mathematics of Computation*, 24, 647-654.