LECTURE NOTE V
Chapter 8
Linearization Methods for Constrained Problems

8.1 Direct use of Successive Linear Programs (SLP)

▶ Linearly Constrained Case

$$\min f(x)$$

subject to $Ax \geq b,\ x \geq 0$

- By linearization of the objective function using 1st-order Taylor series expansion, the lineary constrained problem is converted into LP.

$$f(x) = f(x^0) + \nabla f(x^0)(x - x^0) + O_2 \approx \nabla f(x^0)x + C$$

- *Frank-Wolfe Algorithm*

Given x^0, line search and overall convergence tolerances $\varepsilon > 0$ and $\delta > 0$.

Step 1: Calculate $\nabla f(x^{(k)})$. If $\left\| \nabla f(x^{(k)}) \right\| \leq \varepsilon$, stop. Otherwise, go to step 2.

Step 2: Solve the LP subproblem,

$$\min \nabla f(x^{(k)})y$$

subject to $Ay \geq b,\ y \geq 0$

Let $y^{(k)}$ be the optimal solution to the LP problem.

Step 3: Find $\alpha^{(k)}$ which solve the problem

$$\min f(x^{(k)} + \alpha(y^{(k)} - x^{(k)}))\quad (0 \leq \alpha \leq 1)$$

Step 4: Calculate ($y^{(k)}$ is at the boundary satisfying the constraints.)

$$x^{(k+1)} = x^{(k)} + \alpha^{(k)}(y^{(k)} - x^{(k)})$$

Step 5: Convergence check. If

$$\left\| x^{(k+1)} - x^{(k)} \right\| < \delta \left\| x^{(k+1)} \right\| \text{ and } \left\| f(x^{(k+1)}) - f(x^{(k)}) \right\| < \varepsilon \left\| f(x^{(k+1)}) \right\|$$

then terminate. Otherwise, go to step 1.

Example:

$$\min f(x) = x_1^{0.25} + (x_2/x_1)^{0.25} + (64/x_2)^{0.25}$$

subject to $x_1 \geq 1,\ x_2 \geq x_1,\ 64 \geq x_2$

Suppose as the initial estimate, $x_1=2$ and $x_2=10$ (feasible solution).

$$\frac{\partial f}{\partial x_1} = 0.25 x_1^{-0.75}(1 - x_2^{0.25} x_1^{-0.5}) = -3.83 \times 10^{-2}$$

$$\frac{\partial f}{\partial x_2} = 0.25 x_2^{-0.75}(x_1^{-0.25} - 64^{0.24} x_2^{-0.5}) = -2.38 \times 10^{-3}$$

The 1st subproblem is

$$\min f(x) = -0.0383 y_1 - 0.00238 y_2$$

subject to $y_1 \geq 1,\ y_2 \geq y_1,\ 64 \geq y_2$

and the solution to this LP is $y_1=64$ and $y_2=64$.

The line search of f in the direction $(64\ 64)^T - (2\ 10)^T$ results

$x_1=3.694$ and $x_2=11.475$ ($\alpha = 0.02732$)
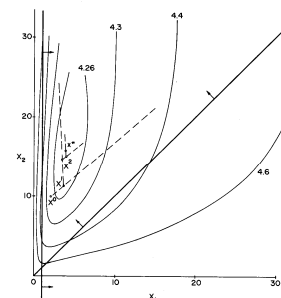
Termination criteria are not met yet.

Figure 8.2. Feasible region, Example 8.1.

Then, the 2nd subproblem becomes

$$\min f(x) = 0.00397\, y_1 - 0.00456\, y_2$$

$$\text{subject to } y_1 \geq 1, \ y_2 \geq y_1, \ 64 \geq y_2$$

and the solution to this LP is $y_1 = 1$ and $y_2 = 64$.

The line search of f in the direction $(1\ 64)^T - (3.694\ 11.475)^T$ results

$$x_1 = 3.526 \text{ and } x_2 = 14.745 \ (\alpha = 0.06225)$$

Through the following iterations, the subproblem solutions approaches to the optimal solution in zigzag pattern.

▶ General NLP Case

$$\min f(x)$$
$$\text{subject to } g(x) \geq 0, \ h(x) = 0, \ x^L \geq x \geq x^U$$

- By linearization of the objective function and constraints using 1st-order Taylor series expansion, the constrained NLP problem is converted into LP.

$$f(x) = f(x^0) + \nabla f(x^0)(x - x^0) + O_2 \approx \nabla f(x^0)x + C$$
$$g(x) = g(x^0) + \nabla g(x^0)(x - x^0) + O_2 \approx \nabla g(x^0)x + b \geq 0$$
$$h(x) = h(x^0) + \nabla h(x^0)(x - x^0) + O_2 \approx \nabla h(x^0)x + a = 0$$

- Not like the linearly constrained case, the obtained solution for LP subproblem is may not be feasible since the constraints of the subproblem is the linearized one.
- Thus, the *direct application* of Frank-Wolfe algorithm does not guarantee any convergence to the optimal solution since the linearization can have large error if the trial point is far from the base point.
- Remedy: limit the region that the LP subproblems can excurse such that the objective function value is decrease and the infeasibility is reduced.

$$-\delta \leq x - x^{(k)} \leq \delta : \text{choose } \delta \text{ so that } f(x^{(k+1)}) - f(x^{(k)}) \leq 0, \ g(x^{(k+1)}) \geq 0$$
$$\text{and } \left| h(x^{(k+1)}) \right| - \left| h(x^{(k)}) \right| \leq 0$$

However, this remedy could slow down the convergence rate quite severely.

- **Use of penalty functions**

Let $P(x, R) = f(x) + R\Omega(g, h, x)$.

Then solve $\min_{\alpha} P(x^{(k)} + \alpha(x^{(k+1)} - x^{(k)}), R)$ in the line search step of the SLP.

Phase Check: If at $x^{(k)}$, the exterior penalty term $\Omega(g, h, x) > \varepsilon$, then go to phase I. Otherwise go to Phase II.

Phase I:

Step 1: Linearize at $x^{(k)}$ and solve the LP subproblem to obtain $\overline{x}$.

Step 2: If $\Omega(g, h, \overline{x}) < \Omega(g, h, x^{(k)})$, then $x^{(k+1)} = \overline{x}$ and proceed to the phase check. Otherwise, reduce δ (e.g., $\delta = \delta / 2$) and go to step 1.

Phase II:

Step 1: Linearize at $x^{(k)}$ and solve the LP subproblem to obtain $\overline{x}$.

Step 2: If $P(\overline{x}, R) < P(x^{(k)}, R)$, set $x^{(k+1)} = \overline{x}$, increase the penalty parameter R, and go to phase check. Otherwise, reduce δ and go to step 1.

Termination: both $\left\| \overline{x} - x^{(k)} \right\|$ and $\left| P(\overline{x}, R) - P(x^{(k)}, R) \right|$ are sufficiently small.

Remark: Still, the convergence results are available only in linearly constrained case.

8.2 Separable Programming

▶ Single-variable separable functions into (K-1) interval

$$\tilde{f}(x) = f^{(k)} + \frac{f^{(k+1)} - f^{(k)}}{x^{(k+1)} - x^{(k)}}(x - x^{(k)}) \quad (\text{for } x^{(k+1)} \geq x \geq x^{(k)})$$

$$x = \lambda^{(k)} x^{(k)} + \lambda^{(k+1)} x^{(k+1)} \quad \text{and} \quad \lambda^{(k)} + \lambda^{(k+1)} = 1$$

$$\tilde{f}(x) = \lambda^{(k)} f^{(k)} + \lambda^{(k+1)} f^{(k+1)}$$

- For the entire range

$$x = \sum_{k=1}^{K} \lambda^{(k)} x^{(k)} \qquad \tilde{f}(x) = \sum_{k=1}^{K} \lambda^{(k)} f^{(k)}$$

where

Figure 8.8. Piecewise linear approximation.

i) $\displaystyle\sum_{k=1}^{K} \lambda^{(k)} = 1$

ii) $\lambda^{(k)} \geq 0 \quad \text{for } k = 1, \cdots, K$

iii) $\lambda^{(i)} \lambda^{(j)} = 0 \quad \text{if } j > i+1 \text{ for } i = 1, \cdots, K-1$

▶ Multivariable separable functions into (K-1) interval

- **Definition**: A function $f(x)$ of N-component vector variable x is said to be *separable* if it can be expressed as the sum of single-variable function that can involve only one of the N variables.

$$f(x) = \sum_{i=1}^{N} f_i(x_i)$$

- **Example**: $f(x) = x_1^2 + e^{x_2} + x_2^{2/3}$ is separable; $f(x) = x_1 \sin(x_2 + x_3) + x_2 e^{x_3}$ is not.

- **Transformation** of non-separable functions:

$$x_1 x_2 \Rightarrow x_3^2 - x_4^2 \quad \text{where} \quad x_3 = 0.5(x_1 + x_2) \quad \text{and} \quad x_4 = 0.5(x_1 - x_2)$$

$$c \prod_{i=1}^{N} x_i^{a_i} \Rightarrow \ln(y) = \ln(c) + \sum_{i=1}^{N} a_i \ln(x_i)$$

- For the entire range

$$\tilde{f}(\mathbf{x}) = \sum_{k=1}^{K_1} \lambda_1^{(k)} f_1^{(k)} + \sum_{k=1}^{K_2} \lambda_2^{(k)} f_2^{(k)} + \cdots + \sum_{k=1}^{K_N} \lambda_N^{(k)} f_N^{(k)}$$

$$x_i = \sum_{k=1}^{K} \lambda_i^{(k)} x_i^{(k)} \quad (i = 1, \cdots, N)$$

where

i) $\displaystyle\sum_{k=1}^{K_i} \lambda_i^{(k)} = 1$

ii) $\lambda_i^{(k)} \geq 0 \quad \text{for } k = 1, \cdots, K_i$

iii) $\lambda_i^{(k)} \lambda_i^{(l)} = 0 \quad \text{if } l > k+1 \text{ for } i = 1, \cdots, K_i - 1$

▶ The NLP can be converted to LP:

$$\min f(x) \approx \sum_{i=1}^{N} \sum_{k=1}^{K_i} \lambda_i^{(k)} f_i^{(k)}$$

$$\text{subject to } g_j(x) \approx \sum_{k=1}^{K_i} \lambda_i^{(k)} g_{ji}^{(k)} \geq 0 \quad (j = 1, \cdots, J)$$

$$\sum_{k=1}^{K_i} \lambda_i^{(k)} = 1 \quad (i = 1, \cdots, N)$$

$$\lambda_i^{(k)} \lambda_i^{(l)} = 0 \quad \text{if } l > k+1 \text{ for } k = 1, \cdots, K_i - 1$$

$$\text{all } \lambda_i^{(k)} \geq 0$$

Remarks: *Restricted basis entry*

1. The difference from normal LP is only the condition for $\lambda_i^{(k)} \lambda_i^{(l)} = 0$.

2. This condition implies that only adjacent $\lambda_i^{(k)}$'s can become basis at the same time for each k. (*Restricted basis entry*)

3. If $\lambda_i^{(k)}$ is going to be a basic, check if there is the entry $\lambda_i^{(k+1)}$ or $\lambda_i^{(k-1)}$ in the basis. If not, choose other variable to become a basic variable.

4. In order to improve the feasibility and accuracy, the number of grid can be increased. However, this increases the number of variable and thus considerably increases the number of iteration due to the restricted basis entry.

5. Or, use coarse grid initially and then use fine grid only in the neighborhood of the resulting solution. However, this may lead to convergence false optima.

6. Use this separable programming only when the nonlinearity is not severe and the objective function and constraints are separable.

- **Example**:

Solve
$$\max f(x) = x_1^4 + x_2$$
$$\text{Subject to } g_1(x) = 9 - 2x_1^2 - 3x_2 \geq 0, \ x_1 \geq 0, \ x_2 \geq 0$$

Sol) $f(x) = f_1(x_1) + f_2(x_2)$, $g(x) = g_{11}(x_1) + g_{12}(x_2)$

where $f_1(x_1) = x_1^4$, $f_2(x_2) = x_2$, $g_{11}(x_1) = -2x_1^2$ and $g_{12}(x_2) = -3x_2$

Division of the range: $0 \leq x_1 \leq 3$ with the interval size=1

$$\tilde{f}_1(x_1) = 0 \cdot \lambda_1^{(1)} + 1 \cdot \lambda_1^{(2)} + 16 \cdot \lambda_1^{(3)} + 81 \cdot \lambda_1^{(4)}$$

$$\tilde{g}_1(x_1) = 0 \cdot \lambda_1^{(1)} - 2 \cdot \lambda_1^{(2)} - 8 \cdot \lambda_1^{(3)} - 18 \cdot \lambda_1^{(4)}$$

Then the linear approximating problem becomes with one slack variable

$$\min f(x) = \lambda_1^{(2)} + 16 \cdot \lambda_1^{(3)} + 81 \cdot \lambda_1^{(4)} + x_2$$

$$\text{Subject to } 2\lambda_1^{(2)} + 8 \cdot \lambda_1^{(3)} + 18 \cdot \lambda_1^{(4)} + 3x_2 + x_3 = 9$$

$$\lambda_1^{(1)} + \lambda_1^{(2)} + \lambda_1^{(3)} + \lambda_1^{(4)} = 1$$

$$\text{all } \lambda_1^{(k)}, x_2, x_3 \geq 0$$

The first tableau is

C_B	Basis	$\lambda_1^{(2)}$	$\lambda_1^{(3)}$	$\lambda_1^{(4)}$	x_2	x_3	$\lambda_1^{(1)}$	b	ratio
0	x_3	2	8	⑱	3	1	0	9	1/2|9/8
0	$\lambda_1^{(1)}$	1	①	1	0	0	1	1	1|1
	r^T	1	16	81	1	0	0	0	

The $\lambda_1^{(4)}$ is a candidate to enter basic and x_3 is leaving the basic. But $\lambda_1^{(1)}$ and $\lambda_1^{(4)}$ cannot

be the basics at the same time (restricted basis entry). Thus, choose $\lambda_1^{(3)}$ and $\lambda_1^{(1)}$ will leave the basis.

C_B	Basis	$\lambda_1^{(2)}$	$\lambda_1^{(3)}$	$\lambda_1^{(4)}$	x_2	x_3	$\lambda_1^{(1)}$	b	ratio
0	x_3	-6	0	⑩	3	1	-8	1	1/10
0	$\lambda_1^{(3)}$	1	1	1	0	0	1	1	1
	r^T	-16	0	65	1	0	-16	16	

The $\lambda_1^{(4)}$ is the candidate for the basis and the restricted basis entry is satisfied.

C_B	Basis	$\lambda_1^{(2)}$	$\lambda_1^{(3)}$	$\lambda_1^{(4)}$	x_2	x_3	$\lambda_1^{(1)}$	b	ratio
0	$\lambda_1^{(4)}$	-0.6	0	1	0.3	0.1	-0.8	0.1	-
0	$\lambda_1^{(3)}$	①.⑥	1	0	-0.3	-0.1	1.8	0.9	5/8
	r^T	23	0	0	-18.5	-6.5	36	97	

The $\lambda_1^{(1)}$ is the candidate for the basis but excluded due to restricted basis entry, then he $\lambda_1^{(2)}$ is the candidate for basic and $\lambda_1^{(3)}$ will leave the basis. This violates the restricted basis entry rule and there is no other choice. Conclude this is the optimal.

$\lambda_1^{(3)}{=}0.9$, $\lambda_1^{(4)}{=}0.1$ ➔ $x_2 =0$, $x_1{=}1x\lambda_1^{(1)}{+} 2x\lambda_1^{(2)}{+} 3x\lambda_1^{(3)}{+} 4x\lambda_1^{(4)}{=}2.1$, $f(x){=}22.5$

(Exact optimum: $x_2 =0$, $x_1{==}2.12$, $f(x){=}20.25$)

Remark: The restricted basis rule will always be satisfied at the optimum if
1. For all i, $f(x_i)$ is either strictly convex or it is linear.
2. For all i and j, $g_{ji}(x_i)$ is either concave or it is linear.

8.3 Cutting Plane Methods

▶ **Motivation**: Concentrate the efforts to construct accurate approximation only when the current point is near optimum.
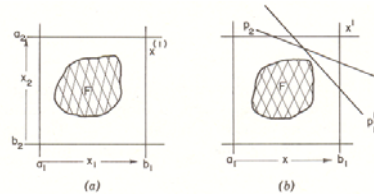

Figure 8.11. (a) Hypercube approximation to feasible region. (b) Cutting planes.

▶ **Kelley's cutting plane method**

min $f(x)$

subject to $g(x)\geq 0$, $x^U \geq x \geq x^L$

- Given a problem in the form of linear objective function (can be obtained from linearization), constraint tolerance $\varepsilon > 0$, and an initial bound of the feasible region (F)

$$Z^0 = \{x : x_i^L \leq x_i \leq x_i^U, i = 1, \cdots, N\} \text{ such that } Z^0 \supset F$$

Step1: Solve linear problem and designate the solution as $x^{(1)}$.

$$\min \sum c_i x_i$$

$$\text{subject to } x \in Z^0$$

For $k=1,2,\ldots$, carry out the following series of steps

Step2: Find m such that (maximum violation)

$$-g_m(x^{(k)}) = \max[-g_j(x^{(k)}), 0; j = 1, \cdots, J]$$

If $g_m(x^{(k)}) > -\varepsilon$, terminate. Otherwise go to step 3.

Step3: Construct the cutting plane,

$$p^{(k)}(x) \equiv \tilde{g}_m(x; x^{(k)}) = g_m(x^{(k)}) + \nabla g_m(x^{(k)})(x - x^{(k)})$$

and let $H^{(k)}$ be the half space $H^{(k)} = \{x : p^{(k)}(x) \geq 0\}$. Solve the LP:

$$\min \sum c_i x_i$$

$$\text{subject to } x \in Z^{(k-1)} \cap H^{(k)}$$

Designate the solution $x^{(k+1)}$.

Step4: Set $Z^{(k)} = Z^{(k-1)} \cap H^{(k)}$ and $k=k+1$. Go to step 2.

- **Example**:

$$\min f(x) = -x_1 - x_2$$

Solve   Subject to $g_1(x) = 2x_1 - x_2^2 - 1 \geq 0$

$$g_2(x) = 9 - 0.8x_1^2 - 2x_2 \geq 0, \ x_1 \geq 0, \ x_2 \geq 0$$

Sol) From the inspection, $Z^0 = \{x : 0 \leq x_1 \leq 5, 0 \leq x_2 \leq 4\}$ is adequately bracket the feasible region for this problem.

First, find the minimum of

$$\min f(x) = -x_1 - x_2$$

$$\text{Subject to } 0 \leq x_1 \leq 5, \ 0 \leq x_2 \leq 4$$

The solution is obviously is $x^{(1)} = (5,4)$.

Since $g_1(x^{(1)}) = -7$ and $g_2(x^{(1)}) = -19$, the most violated constraint is $g_2(x)$.

$$p^{(1)} \equiv g_2(x; x^{(1)}) = -19 - 8(x_1 - 5) - 2(x_2 - 4)$$

The second subproblem becomes:

$$\min f(x) = -x_1 - x_2$$

$$\text{Subject to } 29 - 8x_1 - 2x_2 \geq 0$$

$$0 \leq x_1 \leq 5, \ 0 \leq x_2 \leq 4$$

Figure 8.12. Feasible region, Example 8.11.

The solution is $x^{(2)} = (2.625, 4)$ and $g_1(x^{(2)}) = -11.75$ and $g_2(x^{(2)}) = -4.5125$.

$$p^{(2)} \equiv g_1(x; x^{(2)}) = -11.75 + 2(x_1 - 2.625) - 8(x_2 - 4)$$

The third subproblem becomes:

$$\min f(x) = -x_1 - x_2$$

$$\text{Subject to } 29 - 8x_1 - 2x_2 \geq 0$$

$$15 + 2x_1 - 8x_2 \geq 0$$

$$0 \leq x_1 \leq 5, \ 0 \leq x_2 \leq 4$$

The solution is $x^{(2)} = (2.971, 2.618)$. The iteration continues until the constraints are satisfied within the specified tolerance ε. (The exact optimum is $x^* = (2.5, 2), f^* = -4.5$)

Remarks:

1. The optimum value is approached from below for minimization case since the feasible region is getting smaller. (It generates *sequence of infeasible solutions*)
2. The feasible region F must be convex set.
3. The size of LP problem grows continuously with each iteration.
4. As the solution approaches to the optimum, the constraints become nearly dependent. Thus, needs a procedure to delete old cuts if they are not binding at the optimum of the current subproblem. (*Cut-deletion procedure*)
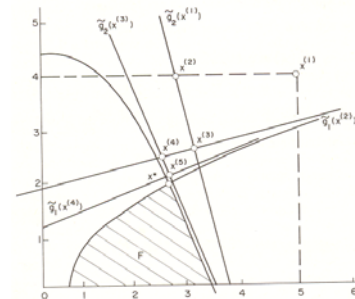
Chapter 9
Direction-Generation Methods Based on Linearization

- Find the decent direction based on the linearized objective function and constraints
- The direction has to be chosen so that it leads to feasible solution.

9.1 Method of Feasible Directions

$$\min f(\mathbf{x})$$

subject to $\mathbf{g}(\mathbf{x}) \geq \mathbf{0}$

The decent direction d while feasible is to satisfy $-\nabla f(x^{(k)})d \geq 0$ and $\nabla g_j(x^{(k)})d \geq 0$.

▶ **Basic Algorithm**

At a given feasible point $x^{(k)}$, let $I^{(k)}$ be the set of indices of those constraints that are active at $x^{(k)}$, within some tolerance $\varepsilon > 0$, that is,

$$I^{(k)} = \{ j : 0 \leq g_j(x^{(k)}) \leq \varepsilon, j = 1, \cdots, J \}$$

Step 1: Solve the LP problem

$$\max \theta$$

$$\text{Subject to } \nabla f(x^{(k)})d \leq -\theta$$

$$\nabla g_j(x^{(k)})d \geq \theta, j \in I^{(k)}$$

$$-1 \leq d_i \leq 1, \; i = 1, \cdots, N$$

Label the solution $d^{(k)}$ and $\theta^{(k)}$.

Step 2: If $\theta^{(k)} \leq 0$, the iteration terminates, since no further improvement is possible. Otherwise, determine

$$\bar{\alpha} = \min\{\alpha \mid g_j(x^{(k)} + \alpha d^{(k)}) = 0, j = 1, \cdots, J \text{ and } \alpha \geq 0\}$$

If no $\bar{\alpha} > 0$ exists, set $\bar{\alpha} = \infty$.

Step 3: Find $\alpha^{(k)}$ such that

$$f(x^{(k)} + \alpha^{(k)}d^{(k)}) = \min\{ f(x^{(k)} + \alpha d^{(k)}), 0 \leq \alpha \leq \bar{\alpha} \}$$

Set $x^{(k+1)} = x^{(k)} + \alpha^{(k)}d^{(k)}$ and continue.

- **Example**:

$$\min f(x) = (x_1 - 3)^2 + (x_2 - 3)^2$$

Solve   Subject to $g_1(x) = 2x_1 - x_2^2 - 1 \geq 0$

$$g_2(x) = 9 - 0.8x_1^2 - 2x_2 \geq 0, \; x_1 \geq 0, \; x_2 \geq 0$$

Sol) The gradients of the problem functions are given by

$$\nabla f = [2(x_1 - 3), 2(x_2 - 3)]$$

$$\nabla g_1 = [2, -2x_2] \text{ and } \nabla g_2 = [-1.6x_1, -2]$$

Suppose the feasible starting point $x^{(k)} = (1,1)$ is given. At this point,

$$g_1(x^{(1)}) = 0.0 \text{ and } g_2(x^{(1)}) = 6.2 > 0$$

Thus, g_1 is the only binding constraint, $I^{(k)} = \{1\}$ and the first subproblem becomes

$$\max \theta$$

$$\text{Subject to } -4d_1 - 4d_2 + \theta \leq 0$$

$$2d_1 - 2d_2 - \theta \geq 0, \quad -1 \leq d_1, d_2 \leq 1$$

The solution to this LP is $d^{(1)}=(1,0)$ with $\theta^{(1)}=2$. We must now search along the ray

$$x = \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \alpha \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1+\alpha \\ 1 \end{pmatrix} \quad \text{and} \quad \alpha \geq 0$$

to find the point at which the boundary of the feasible region is intersected. Since

$$g_1(\alpha) = 2(1+\alpha) - 1 - 1 = 2\alpha, \quad g_2(\alpha) = 9 - 0.8(1+\alpha)^2 - 2$$

the $\bar{\alpha} = 1.958$. The minimum of f with $0 \leq \alpha \leq 1.958$ is

$$f(\alpha) = [(1+\alpha) - 3]^2 + (1-3)^2 = (\alpha - 2)^2 + 4$$

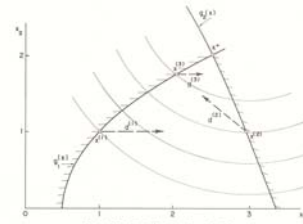Since $\alpha^{(1)} = 1.958$, $x^{(2)}=(2.958,1)$ is obtained.
For the second iteration,

$$\max \theta$$

Subject to $-0.048d_1 - 4d_2 + \theta \leq 0$
$$4.733d_1 - 2d_2 - \theta \geq 0,$$
$$-1 \leq d_1, d_2 \leq 1$$

Figure 9.1. Feasible region, Example 9.1.

The solution to this LP is $d^{(2)}=(-1,0.8028)$ with $\theta^{(2)}=3.127$.
The iterations will continue in the same way.

Remarks:

1. It generates a sequence of feasible solutions.
2. By considering only the active constraints at the current feasible point, a zigzag iteration pattern results, which unfortunately slows down the progress of the iteration.

Figure 9.2. Plot of Example 9.2.

3. It may converge to points that are not KK points in three-dimensional cases. (*jamming*)
4. The remedies are: a) reduce the tolerance for criterion of the active constraints (ε) if $\theta^{(k)}<\varepsilon$.
5. Or, b) replace $\nabla g_j(x^{(k)})d \geq \theta$ with $g_j(x^{(k)}) + \nabla g_j(x^{(k)})d \geq \theta$. The direction will be less affected by the constraint j when $g_j(x^{(k)}) > 0$.
6. The feasible direction methods cannot directly handle the nonlinear equality constraints, because there is no feasible interior to such constraints. To relax this problem, instead of equality, $-\varepsilon \leq h_k \leq \varepsilon$ can be used, but it limits the step size and leads to very slow progress and need a procedure to return to feasible points.
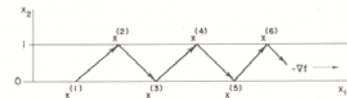
9.2 Simplex Extensions for Linearly constrained Problems

$$\min f(\mathbf{x})$$
$$\text{subject to } \mathbf{Ax} \geq \mathbf{b}$$

Let $\hat{x}$ be the basic variable and $\bar{x}$ be the nonbasic.

$$B\hat{x} + \bar{A}\bar{x} = b \Rightarrow \hat{x} = B^{-1}b, \bar{x} = 0$$

When a linearized objective function is used, all the LP machinery can be used. The *relative cost* can be calculated.

$$r(x^0) = \nabla \bar{f}(x^0) - \nabla \hat{f}(x^0)B^{-1}\bar{A}$$

$$f(x) = f(\hat{x}; \bar{x}) = f(B^{-1}b - B^{-1}\bar{A}\bar{x}, \bar{x})$$

$$f(x) - f(x^0) = r(x^0)(\bar{x} - x^0)$$

For linear case, minimum will be found at the boundary, but for nonlinear case, the minimum may occur before boundary and it requires *line search*.

▶ **Convex Simplex Algorithm**

Given a feasible point x^0, a partition $x = (\hat{x}, \overline{x})$ of the problem variables, and a convergence tolerance $\varepsilon > 0$,

Step 1: Compute $r(x^{(k)})$.

Step 2: Compute β_s and γ_q as follows:

$$\beta_s = \min\{0, r_i : i = 1, \cdots, N - M\}$$

$$\gamma_q = \min\{0, r_i \overline{x}_i : i = 1, \cdots, N - M\}$$

Step 3: If $|\beta_s| \leq \varepsilon$ and $\gamma_q \leq \varepsilon$, terminate. Otherwise, consider two cases:

a) If $|\beta_s| > r_q$, determine $\Delta_r = \min\{\hat{x}_j / p_{js} : p_{js} > 0, j = 1, \cdots, M\}$ and set

$\Delta = \Delta_r$. (where p_{jk} are the elements of matrix $B^{-1}\overline{A}\overline{x}$)

b) If $|\beta_s| < r_q$, determine $\Delta_r = \min\{-\hat{x}_j / p_{jq} : p_{jq} > 0, j = 1, \cdots, M\}$ and set

$\Delta = -\min(\Delta_r, \overline{x}_q)$.

Step 4: Calculate the target point $v^{(k)}$:

$$\hat{v}_j^{(k)} = \hat{x}_j^{(k)} - p_{jk}\Delta \quad \text{and} \quad \overline{v}_j^{(k)} = \begin{cases} \overline{x}_j^{(k)} + \Delta & \text{if } i = k \\ 0 & \text{otherwise} \end{cases}$$

where k is equal to s or q depending upon whether 3(a) or 3(b) occurs.

Step 5: Find α^* such that

$$f(x^{(k)} + \alpha^*(v^{(k)} - x^{(k)})) = \min_{\alpha}\{f(x^{(k)} + \alpha(v^{(k)} - x^{(k)})) : 0 \leq \alpha \leq 1\}$$

Step 6: Set $x^{(k+1)} = x^{(k)} + \alpha^*(v^{(k)} - x^{(k)})$. If $\alpha^* = 1$ and $\Delta = \Delta_r$, update the basis and the basis inverse. Otherwise, retain the same basis and go to step 1.

Remarks:

1. If the optimal solution is not in the corner of the feasible region, the basis will not change for a while. Empirically, occasional change in basis would improve the convergence using periodical random update of basis or choice of basis of *M*-largest variables in magnitude.

2. The rate of convergence to non-corner point optima can be quite slow; however, the direction-finding machinery is very simple and rapid. Thus, one would expect the solution efficiency to be better than that of the Frank-Wolfe algorithm.

▶ **Reduced Gradient Method**

- If all the nonbasic variables change simultaneously instead of changing one-by-one, it would improve the efficiency of the convex simplex method. (more like conjugate gradient)

$$\overline{d}_i = \begin{cases} -r_i & \text{if } r_i \leq 0 \\ -\overline{x}_i r_i & \text{if } r_i > 0 \end{cases} \quad \text{where } i = 1, \cdots, N - M$$

- The gradient of f is reduced to $r(x^0) = \nabla \overline{f}(x^0) - \nabla \hat{f}(x^0) B^{-1}\overline{A}$ in $\overline{x}$ space not in full space of x. (*Reduced gradient*)

- The change in basic variables must be calculated using $\hat{d} = -B^{-1}\overline{A}\overline{d}$ to satisfy the linear constraints.

- For nonbasic variables, $\overline{x}_i$ is increased if the relative cost is negative, while $\overline{x}_i$ is decreased if the relative cost is positive.

Given a feasible point x^0, a partition $x = (\widehat{x}, \overline{x})$ of the problem variables, and a convergence tolerance $\varepsilon > 0$,

Step 1: Compute $r(x^{(k)})$.

Step 2: Compute $\overline{d}$ and $\widehat{d}$ using above relationships. If $\|d\| < \varepsilon$, terminate. Otherwise,

Step 3: Calculate the step-size limit α_{max}=min(α_1, α_2) where

$$\alpha_1 = \min\{-\widehat{x}_i^{(k)} / \widehat{d}_i : \text{for all } \widehat{d}_i < 0, i = 1, \cdots, M\}$$

(If all $\widehat{d}_i \geq 0$, then set $\alpha_1 = \infty$.)

$$\alpha_2 = \min\{-\overline{x}_i^{(k)} / \overline{d}_i : \text{for all } \overline{d}_i < 0, i = 1, \cdots, N-M\}$$

(If all $\overline{d}_i \geq 0$, then set $\alpha_2 = \infty$.)

Step 4: Find α^* such that

$$f(x^{(k)} + \alpha^* d) = \min_{\alpha}\{f(x^{(k)} + \alpha d) : 0 \leq \alpha \leq \alpha_{max}\}$$

Step 5: Calculate the new point $x^{(k+1)} = x^{(k)} + \alpha^* d$. If α^*=α_{max}=α_1, change the basis to avoid degeneracy. Otherwise, go to step 1.

9.3 The Generalized Reduced Gradient Method (GRG)

min $f(\mathbf{x})$

subject to $\mathbf{h}(\mathbf{x}) = \mathbf{0}$ ($i=1,\dots,I$)

▶ **Implicit Variable Elimination**

- Reduce the number of variable to search using the equality constraints

- Using the linearized equality at feasible point $x^{(k)}$ and for $\widetilde{h}_i(x; x^{(k)})$ to be feasible,

$$\widetilde{h}_i(x; x^{(k)}) \equiv h_i(x^{(k)}) + \nabla h_i(x^{(k)})(x - x^{(k)}) \quad (i = 1, \cdots, I)$$

$$\widetilde{h}_k(x; x^{(k)}) = 0 \Rightarrow \nabla h_k(x^{(k)})(x - x^{(k)}) = 0 \quad (k = 1, \cdots, K)$$

- Using the partition $x = (\widehat{x}, \overline{x})$

$$\mathbf{J}(\widehat{x} - \widehat{x}^{(k)}) + \mathbf{C}(\overline{x} - \overline{x}^{(k)}) = 0$$

where $\mathbf{J} = (\nabla_1 \widehat{h} \ \nabla_2 \widehat{h} \ \cdots \nabla_I \widehat{h})^T$ and $\mathbf{C} = (\nabla_1 \overline{h} \ \nabla_2 \overline{h} \ \cdots \nabla_{N-I} \overline{h})^T$

$$(\widehat{x} - \widehat{x}^{(k)}) = -\mathbf{J}^{-1}\mathbf{C}(\overline{x} - \overline{x}^{(k)})$$

- Eliminate the variable $\widehat{x}$

$$\widetilde{f}(\widehat{x}; \overline{x}) \cong f(\widehat{x}^{(k)} - \mathbf{J}^{-1}\mathbf{C}(\overline{x} - \overline{x}^{(k)}), \overline{x}) \quad \text{(a function of } N\text{–}I \text{ variables)}$$

- Optimality conditions for $\widetilde{f}(\overline{x}) = \widetilde{f}(\widehat{x}(\overline{x}), \overline{x})$

$$\frac{\partial \widetilde{f}}{\partial \overline{x}} = \frac{\partial f}{\partial \overline{x}} + \frac{\partial \widetilde{f}}{\partial \widehat{x}}\frac{\partial \widehat{x}}{\partial \overline{x}} \Rightarrow \nabla \widetilde{f}(x^{(k)}) = \nabla \overline{f}(x^{(k)}) + \nabla \widehat{f}(x^{(k)})\mathbf{J}^{-1}\mathbf{C} = 0$$

cf) This condition is equivalent to Lagrangian stationary condition:

$$\nabla f(x^*) - v^* \nabla h(x^*) = 0 \Rightarrow \begin{cases} \nabla \widehat{f}(x^*) - v^{*T}\mathbf{J} = 0 \\ \nabla \overline{f}(x^*) - v^{*T}\mathbf{C} = 0 \end{cases}$$

$$v^* = (\nabla \widehat{f}(x^*)\mathbf{J}^{-1})^T \Rightarrow \nabla \overline{f}(x^*) - \nabla \widehat{f}(x^*)\mathbf{J}^{-1}\mathbf{C} = 0$$

▶ **Generalized Reduced Gradient Algorithm**

Given a feasible point x^0, a partition $x = (\hat{x}, \bar{x})$ which has associated with it a constraint gradient submatrix **J** with nonzero determinant, a specified initial value of search parameter $\alpha = \alpha^0$, termination parameter ε_1, ε_2 and ε_3, all positive, and a reduction parameter γ, $0 < \gamma < 1$,

Step 1: Calculate $\nabla \tilde{f} = \nabla \bar{f}(x^{(k)}) - \nabla \hat{f}(x^{(k)}) \mathbf{J}^{-1} \mathbf{C}$.

Step 2: If $\left\| \nabla \tilde{f} \right\| \leq \varepsilon$, stop. Otherwise, set $d = (\hat{d}, \bar{d})^T$ where

$$\bar{d} = (-\nabla \tilde{f})^T \quad \text{and} \quad \hat{d} = -\mathbf{J}^{-1} \mathbf{C} \bar{d}$$

Step 3: (minimizing f in d direction while satisfying the equality constraints)

Set $\alpha = \alpha^0$, For $l = 1, 2, \ldots$,

a) Calculate $v^{(k)} = x^{(k)} + \alpha d$. If $\left| h_i(v^{(k)}) \right| \leq \varepsilon_2$ for $i = 1, \ldots, I$, go to (d). Otherwise, continue.

b) Let $\hat{v}^{(l+1)} = \hat{v}^{(l)} - \mathbf{J}^{-1} v^{(l)} h(v^{(l)})$ and $\bar{v}^{(l+1)} = \bar{v}^{(l)}$.

c) If $\left\| \hat{v}^{(l+1)} - \hat{v}^{(l)} \right\| > \varepsilon_3$, go to b). Otherwise, if $\left| h_i(v^{(l)}) \right| \leq \varepsilon_2$ for $i = 1, \ldots, I$, go to (d). Otherwise, set $\alpha = \gamma \alpha$ and go to (a).

d) If $f(x^{(k)}) \leq f(v^{(l)})$, set $\alpha = \gamma \alpha$ and go to (a). Otherwise, set $x^{(k+1)} = v^{(l)}$ and go to step 1.

- **Example**:

$$\min f(x) = 4x_1 - x_2^2 + x_3^2 - 12$$

Solve   Subject to $h_1(x) = 20 - x_1^2 - x_2^2 = 0$

$$h_2(x) = x_1 + x_3 - 7 = 0$$

Sol) Suppose we are given the feasible starting point $x^{(1)} = (2,4,5)$ and suppose we choose x_1 as nonbasic variable and x_2 and x_3 as basic variables. Thus,

$$\hat{x} = (x_2, x_3) \quad \text{and} \quad \bar{x} = x_1$$

The function derivatives are

$$\nabla f = (4, -2x_2, 3x_3), \quad \nabla h_1 = (-2x_1, -2x_2, 0) \quad \text{and} \quad \nabla h_2 = (1, 0, 1).$$

Step 1: Since $\nabla f = (4, -8, 10)$, $\nabla h_1 = (-4, -8, 0)$ and $\nabla h_2 = (1, 0, 1)$,

$$\mathbf{J} = \begin{pmatrix} -8 & 0 \\ 0 & 1 \end{pmatrix} \quad \mathbf{C} = \begin{pmatrix} -4 \\ 1 \end{pmatrix} \quad \nabla \hat{f} = (-8, 10) \quad \nabla \bar{f} = (4)$$

$$\mathbf{J}^{-1} = \begin{pmatrix} -1/8 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad \nabla \tilde{f} = (4) - (-8, 10) \begin{pmatrix} -1/8 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} -4 \\ 1 \end{pmatrix} = -2$$

Step 2: The direction vector becomes

$$\bar{d} = -\nabla \tilde{f} = 2 \quad \text{and} \quad \hat{d} = -\begin{pmatrix} -1/8 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} -4 \\ 1 \end{pmatrix} 2 = \begin{pmatrix} -1 \\ -2 \end{pmatrix}$$

Step 3: Set $\alpha^0 = 1$.

a) $v^{(1)T} = (2,4,5) + \alpha(2,-1,-2) = (4,3,3)$, $h_1(v^{(1)}) = -5$ and $h_2(v^{(1)}) = 0$. (not small enough)

b) $\hat{v}^{(2)} = \begin{pmatrix} 3 \\ 3 \end{pmatrix} - \begin{pmatrix} -6 & 0 \\ 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} -5 \\ 0 \end{pmatrix} = \begin{pmatrix} 2.167 \\ 3 \end{pmatrix}$

c) Since $\left\| \hat{v}^{(2)} - \hat{v}^{(1)} \right\|$ is not small enough, repeat b) again.

b) $\widehat{v}^{(3)} = \begin{pmatrix} 13/6 \\ 3 \end{pmatrix} - \begin{pmatrix} -13/3 & 0 \\ 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} -0.69 \\ 0 \end{pmatrix} = \begin{pmatrix} 2.008 \\ 3 \end{pmatrix}$

c) Assume this result is a converged one. Checking $h_1(v^{(3)})= -0.032$ and $h_2(v^{(3)})=0$ is assumed to pass the condition.

d) $v^{(3)T}=(4,2.008,3)$ and $f(x^{(1)})=5< f(v^{(3)})=9$: No improvement in $f(x)$.

Reduce a as α=0.4 (γ=0.4) and repeat step 3.

a) $v^{(4)T}=(2,4,5)+ (0.4)(2,-1,-2)=(2.8,3.6,4.2)$, $h_1(v^{(1)})= -0.8$ and $h_2(v^{(1)})=0$. (not small enough, Newton iteration b)~c) is required.)

b) $\widehat{v}^{(5)} = \begin{pmatrix} 3.6 \\ 4.2 \end{pmatrix} - \begin{pmatrix} -7.2 & 0 \\ 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} -0.8 \\ 0 \end{pmatrix} = \begin{pmatrix} 3.49 \\ 4.2 \end{pmatrix}$

c) Assume $\left\| \widehat{v}^{(5)} - \widehat{v}^{(4)} \right\|$ is small enough, and $h_1(v^{(5)})= -0.02$ and $h_2(v^{(5)})=0$ is assumed to pass the condition.

d) $v^{(5)T}=(2.8,3.29,4.2)$ and $f(x^{(1)})=5< f(v^{(5)})=4.66$: Improved!

Set $x^{(2)}= v^{(5)}= (2.8,3.29,4.2)^T$ and repeat from step 1.

(Exact optimum: $x^{*}= (2.5,3.71,4.5)^T$)

Remarks:
1. The Newton iteration for feasibility is the most time-consuming step.
2. Stop Newton iteration if there is improvement.
3. ***Treatment of bounds***: Treating them as inequality is not very good idea.
   a) A check must be made to ensure that only variables that are not on or very near their bounds are labeled as basic variables. This check is necessary to ensure that some free adjustment of the basic variables can always be undertaken
   ➔ Order the variable in the magnitude of the distance from their bounds and choose largest I variables such that the **J** becomes nonsingular.
   b) The direction $\overline{d}$ is modified to ensure that the bounds on the independent variables will not be violated if movement is undertaken in the $\overline{d}$ direction. This is accomplished by setting

$$\overline{d}_i = \begin{cases} 0 & \text{if } \overline{x}_i = \overline{x}_i^{(U)} \text{ and } \nabla f_i <0, \\ 0 & \text{if } \overline{x}_i = \overline{x}_i^{(L)} \text{ and } \nabla f_i >0, \\ -\nabla f_i & \text{otherwise} \end{cases}$$

   c) The checks must be inserted in step 3 of the GRG algorithm to ensure that the bounds are not exceed during the Newton iterations.
4. ***Treatment of inequalities***: If the inequality is given by $a_j \le g_j(x) \le b_j$,

   use $h_{I+j}(x) = g_j(x) - x_{N+j} = 0$ with $a_j \le x_{N+j} \le b_j$

   ➔ Regard them as equality and bounds on new slack variables.
5. ***Treatment of linear elements***: Each problem functions can be subdivided into linear and nonlinear components for linear and nonlinear variables, y and x.

   The NLP problem can be expresses as follow:

$$\min f(x)+c^T x+d^T y$$
$$\text{subject to } h(x)+A_1 y = b_1$$
$$A_2 x+A_3 y = b_2$$
$$l \le x, y \le u$$

   The problem variables are divided into three sets of variables

- The *I* basic variables which have values strictly between their bounds.
- The S superbasic variables which have values strictly between their bounds but are not basic.
- The *N–I–S* nonbasic variables which have values lie on one or the other of the bounds.
- The nonlinear part will be linearized. Then, the equality becomes (B is chosen so that it is nonsingular)

    $Bx_B + Sx_S + Nx_N = 0$ ➜ $x_B + B^{-1}Sx_S + B^{-1}Nx_N = 0$

- At a solution, the basic and superbasic variables will lie somewhere between their bounds (to within the feasibility tolerance), while nonbasic variables will normally be equal to one of their bounds.
- At a solution, the number of superbasic variables is no more than the number of nonlinear variables.
- In the reduced gradient algorithms, x_S will be regarded as a set of independent variables or free variables that are allowed to move in any desirable direction to improve the objective function value or to reduce the sum of infeasibilities. The basic variables can then be adjusted to satisfy the constraints.
- If it appears that no improvement can be made with current definitions of B, S, and N, some of the nonbasic variables are selected to be added to S with an increased size of S (number of columns). At all stages, if a basic or superbasic variable encounters one of its bound, the variable is made nonbasic and the size of S is reduced by 1.
- The number of superbasic variables indicates the number of degrees of freedom remaining after the constraints have been satisfied. In broad terms, the number of superbasic variables is *a measure of how nonlinear the problem is*.

6. MINOS
   - Basically GRG algorithm without the *restoration* of the equality constraints.
   - Instead, use augmented Lagrange objective function

   $$Q(x) = f(x) + v^T h(x) + \beta h(x)^T h(x)$$