INEQUALITY CONSTRAINTS

Introduction of Slack Variables

- Consider the very general situation in which we have a nonlinear objective function, nonlinear equality, and nonlinear inequality constraints.

- The simplest way to handle inequality constraints is to convert them to equality constraints using ***slack variables*** and then use the Lagrange theory.

- Consider the inequality constraints

$$h_j(\boldsymbol{x}) \geq 0 \qquad j = 1, 2 \ldots, r$$

  and define the real-valued slack variables θ_j such that

$$\theta_j^2 = h_j(\boldsymbol{x}) \geq 0 \qquad j = 1, 2, \ldots, r$$

  but at the expense of introducing r new variables.

- If we now consider the general problem written as

$$\underset{\boldsymbol{x}}{\text{minimize}} \qquad f(\boldsymbol{x}) \tag{1}$$

$$\text{subject to} \qquad h_j(\boldsymbol{x}) \geq 0 \quad j = 1(1)r \tag{2}$$

- Introducing the slack variables:

$$h_j(\boldsymbol{x}) - \theta_j^2 = 0 \qquad j = 1(1)r$$

  the Lagrangian is written as:

$$L(\boldsymbol{x}, \lambda, \theta) = f(\boldsymbol{x}) + \sum_{j=1}^{r} \lambda_j (h_j(\boldsymbol{x}) - \theta_j^2) \qquad (3)$$

- The *necessary conditions* for an optimum are:

$$\frac{\partial L}{\partial x_i} = \frac{\partial f}{\partial x_i} + \sum_{j=1}^{r} \lambda_j \frac{\partial h_j}{\partial x_i}\Bigg|_{\substack{\boldsymbol{x} = \boldsymbol{x}^* \\ \lambda = \lambda^*}} = 0 \qquad i = 1(1)n \qquad (4)$$

$$\frac{\partial L}{\partial \lambda_j} = h_j(\boldsymbol{x}) - \theta_j^2 \Bigg|_{\substack{\boldsymbol{x} = \boldsymbol{x}^* \\ \theta = \theta^*}} = 0 \qquad j = 1(1)r \qquad (5)$$

$$\frac{\partial L}{\partial \theta_j} = -2\lambda_j^* \theta_j^* = 0 \qquad j = 1(1)r \qquad (6)$$

- From the last expression (6), it is obvious that either $\lambda^* = 0$ or $\theta_j^* = 0$ or both.

- <u>Case 1</u>: $\lambda_j^* = 0$, $\theta_j^* \neq 0$

In this case, the constraint $h_j(\boldsymbol{x}) \geq 0$ is ignored since $h_j(\boldsymbol{x}^*) = (\theta_j^*)^2 > 0$ (*i.e.* the constraint is not binding).

If all $\lambda_j^* = 0$, then (4) implies that $\nabla f(\boldsymbol{x}^*) = 0$ which means that the the solution is the unconstrained minimum.

- Case 2: $\theta_j^* = 0$, $\lambda_j^* \neq 0$

  In this case, we have $h_j(\boldsymbol{x}^*) = 0$ which means that the optimal solution is on the boundary of the j^{th} constraint.

  Since $\lambda_j^* \neq 0$ this implies that $\nabla f(\boldsymbol{x}^*) \neq 0$ and therefore we are not at the unconstrained minimum.

- Case 3: $\theta_j^* = 0$ and $\lambda_j^* = 0$ for all j.

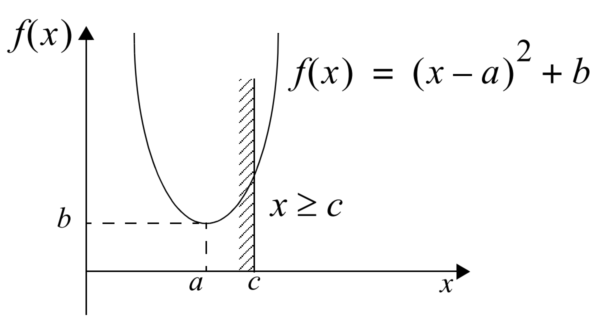
  In this case, we have $h_j(\boldsymbol{x}^*) = 0$ for all j and $\nabla f(\boldsymbol{x}^*) = 0$.

  Therefore, the boundary passes through the unconstrained optimum which is also the constrained optimum.

Example: Now consider the problem

$$\underset{x}{\text{minimize}} \qquad f(x) = (x-a)^2 + b$$

$$\text{subject to:} \qquad x \geq c$$



Sketch of the constrained one dimensional problem.

- The location of the minimum depends on whether or not the unconstrained minimum is inside the feasible region or not.

- If $c > a$ then the minimum lies at $x = c$, which is the boundary of the feasible reagion defined by $x \geq c$.

- If $c \leq a$ then the minimum lies at the unconstrained minimum, $x = a$.

- Introducing a single slack variable, $\theta^2 = x - c \geq 0$:

$$x - c - \theta^2 = 0$$

and we can write the Lagrangian as

$$L(x, \lambda, \theta) = (x - a)^2 + b + \lambda(x - c - \theta^2)$$

where λ is the Lagrange multiplier.

$$\frac{\partial L}{\partial x} = 2(x^* - a) + \lambda^* = 0 \tag{7}$$

$$\frac{\partial L}{\partial \lambda} = x^* - c - \theta^{*2} = 0 \tag{8}$$

$$\frac{\partial L}{\partial \theta} = -2\lambda^* \theta^* = 0 \tag{9}$$

- In general, we need to know how c and a compare.

- Case 1:

  From (9), assume $\lambda^* = 0$, $\theta^* \neq 0$.

  Therefore, from (7) $x^* = a$ and thus from (8), $a - c - \theta^{*2} = 0$ which gives that $\theta^{*2} = a - c$ and we have that θ^* is real only for $c \leq a$.

  Now since $\lambda^* = 0$ we have

  $$L(x^*, \lambda^*, \theta^*) = f(x^*) \text{ and } \left.\frac{\partial f}{\partial x}\right|_{x^*} = 0$$

  This tells us that the unconstrained minimum is the constrained minimum.

- Case 2: Now let us assume that $\lambda^* \neq 0$, $\theta^* = 0$.

  From (8) we have $x^* = c$ and from (7) $\lambda^* = 2(a - c)$.

  Since $\lambda^* \neq 0$ and in the previous case we had $c \leq a$, now we have $c > a$.

- Case 3: For the case $\lambda^* = \theta^* = 0$, (7) tells us that $x^* - a = 0$ and therefore $x^* = a$.

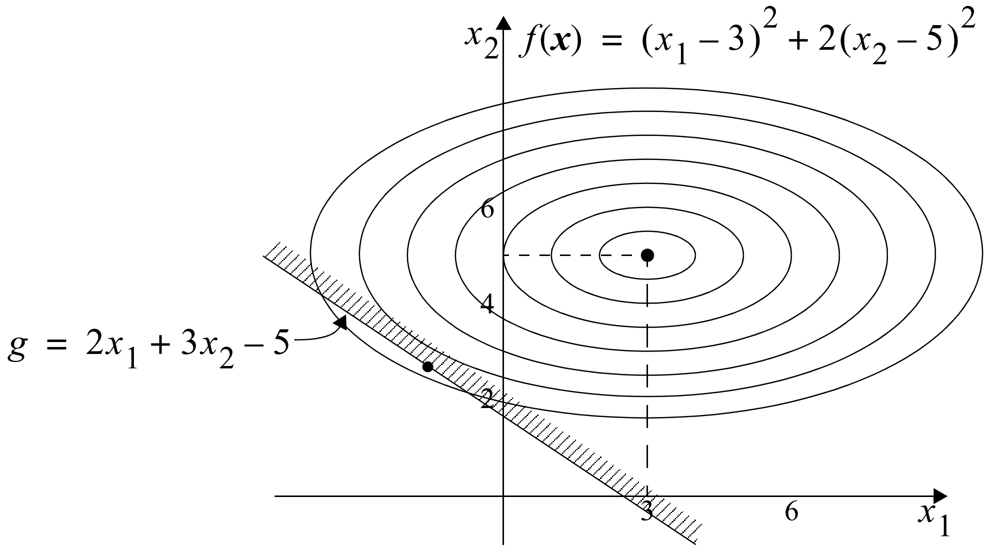
  From (8) we have $x - c = 0$ and therefore $x^* = a = c$. The uncostrained minimum lies on the boundary since from (7)

  $$\left.\frac{\partial L}{\partial x}\right|_{x^*} = \left.\frac{\partial f}{\partial x}\right|_{x^*} = 0.$$

Example: As a two dimensional example, consider

$$\underset{\boldsymbol{x}}{\text{minimize}} \quad f(\boldsymbol{x}) = (x_1 - 3)^2 + 2(x_2 - 5)^2$$

$$\text{subject to:} \quad g(\boldsymbol{x}) = 2x_1 + 3x_2 - 5 \le 0$$



Contours and feasible region for the example problem.

- Unless one was to draw a very accurate contour plot, it is hard to find the minimum from such a graphical method.

- It is obvious from the graph though, that the minimum will lie on the line $g(\boldsymbol{x}) = 0$.

- We introduce a single slack variable, θ^2, and construct the Lagrangian as

$$L(\boldsymbol{x}, \lambda, \theta) = (x_1 - 3)^2 + 2(x_2 - 5)^2 + \lambda(2x_1 + 3x_2 - 5 + \theta^2).$$

- The inequality constraint was changed to the equality constraint $g(\boldsymbol{x}) + \theta^2 = 0$, using the slack variable $\theta^2 = -g(\boldsymbol{x}) \ge 0$.

- The necessary conditions become

$$\frac{\partial L}{\partial x_1} = 2(x_1^* - 3) + 2\lambda^* = 0 \tag{10}$$

$$\frac{\partial L}{\partial x_2} = 4(x_2^* - 5) + 3\lambda^* = 0 \tag{11}$$

$$\frac{\partial L}{\partial \theta} = 2\theta^* \lambda^* = 0 \tag{12}$$

$$\frac{\partial L}{\partial \lambda} = 2x_1^* + 3x_2^* - 5 + \theta^{*2} = 0 \tag{13}$$

From (10) and (11):

$$x_1^* = 3 - \lambda^*$$

$$x_2^* = 5 - \frac{3}{4}\lambda^*$$

substituting these expressions in (13) we have:

$$2(3 - \lambda^*) + 3\left(5 - \frac{3}{4}\lambda^*\right) - 5 + \theta^{*2} = 0$$

$$16 - \frac{17}{4}\lambda^* + \theta^{*2} = 0.$$

If $\lambda^* = 0$ then θ^* will be complex. If $\theta^* = 0$ then $\lambda^* = 64/17$ and therefore

$$x_1^* = -\frac{13}{17} \qquad x_2^* = \frac{37}{17}$$

$\theta^* = 0$ means there is no slack in the constraint as expected from the plot.

The Kuhn-Tucker Theorem

- Kuhn-Tucker theorem gives the necessary conditions for optimum of a nonlinear objective function constrained by a set of nonlinear inequality constraints.

- The general problem is written as

$$\underset{\boldsymbol{x}}{\text{minimize}} \quad f(\boldsymbol{x}) \qquad x \in \subsetneq^n$$

$$\text{subject to:} \quad g_i(\boldsymbol{x}) \le 0 \qquad i = 1, 2, \ldots r$$

If we had equality constraints, then we could introduce two inequality constraints in place of it.

For instance if it was required that $h(\boldsymbol{x}) = 0$, then we could just impose $h(\boldsymbol{x}) \le 0$ and $h(\boldsymbol{x}) \ge 0$ or $-h(\boldsymbol{x}) \le 0$.

- Now assume that $f(\boldsymbol{x})$ and $g_i(\boldsymbol{x})$ are differentiable functions; The Lagrangian is:

$$L(\boldsymbol{x}, \lambda) = f(\boldsymbol{x}) + \sum_{i=1}^{r} \lambda_i g_i(\boldsymbol{x})$$

The necessary conditions for $\boldsymbol{x}^*$ to be the solution to the above problem are:

$$\frac{\partial}{\partial x_j} f(\boldsymbol{x}^*) + \sum_{i=1}^{r} \lambda_i^* \frac{\partial}{\partial x_j} g_i(\boldsymbol{x}^*) = 0 \quad j = 1, 2, \ldots, n \tag{14}$$

$$g_i(\boldsymbol{x}^*) \le 0 \qquad i = 1(1)r \tag{15}$$

$$\lambda_i^* g_i(\boldsymbol{x}^*) = 0 \qquad i = 1(1)r \tag{16}$$

$$\lambda_i^* \ge 0 \qquad i = 1(1)r \tag{17}$$

- These are known as the Kuhn-Tucker stationary conditions; written compactly as:

$$\nabla_{\boldsymbol{x}} L(\boldsymbol{x}^*, \lambda^*) = \mathbf{0} \tag{18}$$

$$\nabla_{\lambda} L(\boldsymbol{x}^*, \lambda^*) = \boldsymbol{g}(\boldsymbol{x}^*) \le \mathbf{0} \tag{19}$$

$$(\lambda^*)^T \boldsymbol{g}(\boldsymbol{x}^*) = \mathbf{0} \tag{20}$$

$$\lambda^* \ge \mathbf{0} \tag{21}$$

- If our problem is one of maximization instead of minimization then

$$\begin{array}{lll} \underset{\boldsymbol{x}}{\text{maximize}} & f(\boldsymbol{x}) & x \in \subsetneq^n \\ \text{subject to:} & g_i(\boldsymbol{x}) \le 0 & i = 1, 2, \ldots r \end{array}$$

we can replace $f(\boldsymbol{x})$ by $-f(\boldsymbol{x})$ in the first condition

$$-\frac{\partial}{\partial x_j} f(\boldsymbol{x}^*) + \sum_{i=1}^{r} \lambda_i^* \frac{\partial}{\partial x_j} g_i(\boldsymbol{x}^*) = 0 \qquad j = 1, 2, \ldots, n \tag{22}$$

$$\frac{\partial}{\partial x_j} f(\boldsymbol{x}^*) + \sum_{i=1}^{r} (-\lambda_i^*) \frac{\partial}{\partial x_j} g_i(\boldsymbol{x}^*) = 0 \quad j = 1, 2, \ldots, n. \tag{23}$$

- For the maximization problem is one of changing the sign of λ_i^* :

$$\nabla_{\boldsymbol{x}} L(\boldsymbol{x}^*, \lambda^*) = \mathbf{0} \tag{24}$$

$$\nabla_{\lambda} L(\boldsymbol{x}^*, \lambda^*) = \boldsymbol{g}(\boldsymbol{x}^*) \le \mathbf{0} \tag{25}$$

$$(\lambda^*)^T \boldsymbol{g}(\boldsymbol{x}^*) = \mathbf{0} \tag{26}$$

$$\lambda^* \leq \mathbf{0} \qquad (27)$$

Transformation via the Penalty Method

- The Kuhn-Tucker necessary conditions give us a theoretical framework for dealing with nonlinear optimization

- From a practical computer algorithm point of view we are not much further than we were when we started.

- We require practical methods of solving problems of the form:

$$\underset{\boldsymbol{x}}{\text{minimize}} \quad f(\boldsymbol{x}) \qquad \boldsymbol{x} \in \subsetneq^n \tag{28}$$

$$\text{subject to} \quad g_j(\boldsymbol{x}) \leq 0 \qquad j = 1(1)J \tag{29}$$

$$h_k(\boldsymbol{x}) = 0 \qquad k = 1(1)K \tag{30}$$

- We introduce a new objective function called the ***penalty function***

$$P(\boldsymbol{x};\boldsymbol{R}) = f(\boldsymbol{x}) + \Omega(\boldsymbol{R}, \boldsymbol{g}(\boldsymbol{x}), \boldsymbol{h}(\boldsymbol{x}))$$

  where the vector $\boldsymbol{R}$ contains the penalty parameters and

$$\Omega(\boldsymbol{R}, \boldsymbol{g}(\boldsymbol{x}), \boldsymbol{h}(\boldsymbol{x}))$$

  is the penalty term.

- The penalty term is a function of $\boldsymbol{R}$ and the constraint functions, $\boldsymbol{g}(\boldsymbol{x}), \boldsymbol{h}(\boldsymbol{x})$.

- The purpose of the addition of this term to the objective function is to **penalize** the objective function when a set of decision variables, $\boldsymbol{x}$, which are not feasible are chosen.

Use of a parabolic penalty term

- Consider the minimization of an objective function, $f(\boldsymbol{x})$ with equality constraints, $\boldsymbol{h}(\boldsymbol{x})$.

- We create a penalty function by adding a positive coefficient times each constraint, that is

$$\underset{\boldsymbol{x}}{\text{minimize}} \quad P(\boldsymbol{x};\boldsymbol{R}) = f(\boldsymbol{x}) + \sum_{k=1}^{K} R_k \{h_k(x)\}^2 . \qquad (31)$$

As the penalty parameters $R_k \to \infty$, more weight is attached to satisfying the k^{th} constraint.

If a specific parameter is chosen as zero, say $R_k = 0$, then the k^{th} equality constraint is ignored.

The user specifies value of R_k according to the importance of satisfying each equality constraint.

Example:

$$\underset{\boldsymbol{x}}{\text{minimize}} \quad x_1^2 + x_2^2$$

$$\text{subject to:} \quad x_2 = 1$$

We construct a penalty function as:

$$P(\boldsymbol{x};R) = x_1^2 + x_2^2 + R(x_2 - 1)^2$$

and we proceed to minimizing $P(\boldsymbol{x};R)$ for particular values of R.

- We proceed analytically; first order necessary conditions for a minimum

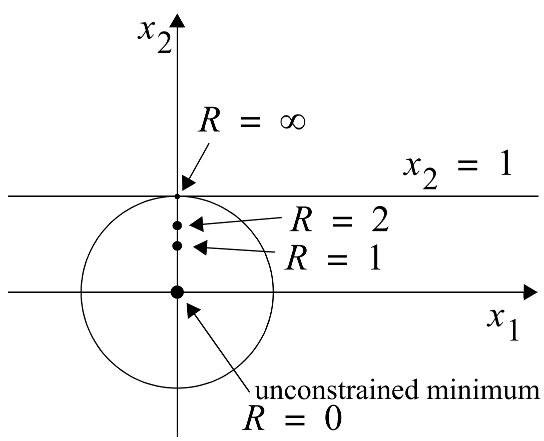
$$\frac{\partial P}{\partial x_1} = 2x_1^* = 0 \qquad \Rightarrow x_1^* = 0$$

$$\frac{\partial P}{\partial x_2} = 2x_2^* + 2R(x_2^* - 1) = 0 \qquad \Rightarrow x_2^* = \frac{R}{1+R}$$

If we now take the limit as $R \to \infty$, we have

$$x_2^* = \lim_{R \to \infty} \frac{R}{1+R} = 1 .$$

- In a numerical procedure, the value of the R would be increased gradually and the numerical optimization would be performed several times.



Example of the use of a parabolic penalty function.

Inequality constrained problems

- Consider the penalty method for inequality constrained problems.

- The general nonlinear objective function with J nonlinear inequality constraints is written as

$$\underset{\boldsymbol{x}}{\text{minimize}} \quad f(\boldsymbol{x}) \qquad \boldsymbol{x} \in \mathbb{R}^n \tag{32}$$

$$\text{subject to} \quad g_j(\boldsymbol{x}) \le 0 \qquad j = 1(1)J \tag{33}$$

A penalty function can be constructed as

$$P(\boldsymbol{x};R) = f(\boldsymbol{x}) + \sum_{i=1}^{J} R_i [g_i(x)]^2 u(g_i) \tag{34}$$

where $u(g_i)$ is the step-function defined by

$$u(g_i) = \begin{cases} 0 & \text{if } g_i(x) \le 0 \\ 1 & \text{if } g_i(x) > 0 \end{cases} \tag{35}$$

and the penalty parameter is chosen as a positive number, $R_i > 0$.

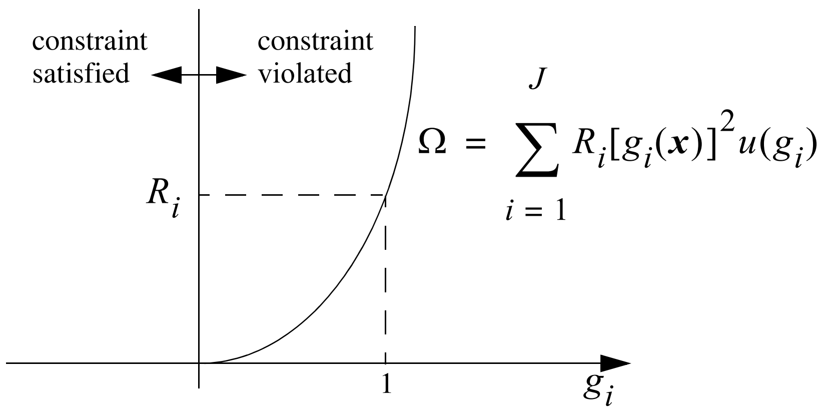
- The term $[g_i(x)]^2 u(g_i)$ is sometimes called the *bracket operator* and is denoted

$$\langle g_i(x) \rangle$$

- The step-function is used to ignore the constraint when it is satisfied by the decision variables and to treat it as a penalty term when it is not satisfied.

- When this type of penalty term is used, the method is referred to as an *exterior penalty method* since points outside the feasible region are allowed, but are penalized.

- As the penalty parameter increases, the feasibility region is "*pushed in*".



Exterior penalty method.

Inverse penalty term

- An alternate method which is commonly used is the inverse penaly method.

- If we have a nonlinear optimization problem written as

$$\underset{\boldsymbol{x}}{\text{minimize}} \quad f(\boldsymbol{x}) \qquad \boldsymbol{x} \in \mathbb{R}^n \tag{36}$$

$$\text{subject to} \quad g_j(\boldsymbol{x}) \geq 0 \qquad j = 1(1)J \tag{37}$$

then we can construct a penalty function as

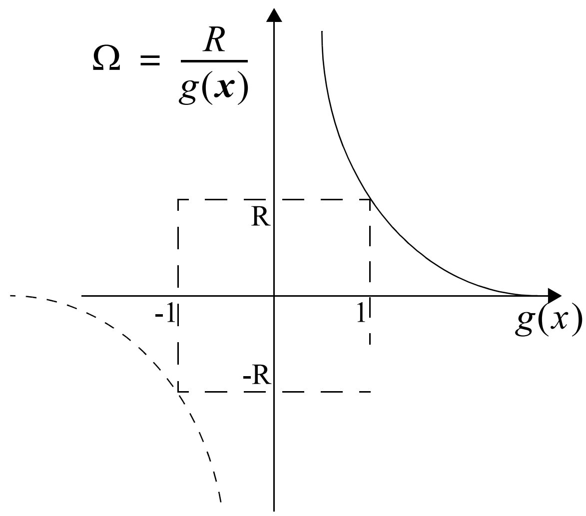
$$P(\boldsymbol{x};R) = f(\boldsymbol{x}) + R \sum_{i=1}^{J} \frac{1}{g_i(x)}$$

where only **one** penalty parameter is used.

- The method is easy to visualize if we consider the case of only one constraint, $J = 1$. Then the penalty term is simply

$$\Omega = \frac{R}{g(\boldsymbol{x})}$$

where $g(\boldsymbol{x})$ is the single constraint.



Inverse penalty term.

- As can be deduced from the figure, it is important that only feasible points be started with; because of this, this method is classified as an *interior method*.

- With *exterior penalties* the parameter R is steadily increased with $R \to \infty$ in the limit so as to exclude infeasible points.

- With *interior penalties*, the parameter R is steadily decreased with $R \to 0$ in the limit. Otherwise, you may artificially exclude a minimum located on the boundary.

Example:

In the following problem, solve it using both an interior and exterior method.

$$\underset{\boldsymbol{x}}{\text{minimize}} \qquad f(\boldsymbol{x}) = (x_1 - 4)^2 + (x_2 - 4)^2$$

$$\text{subject to:} \qquad g(\boldsymbol{x}) = 5 - x_1 - x_2 \geq 0$$

solve using the "*bracket operator*"

$$P(\boldsymbol{x};R) = f(\boldsymbol{x}) + R[g(x)]^2 u(-g(\boldsymbol{x}))$$

we have

$$P(\boldsymbol{x};R) = (x_1 - 4)^2 + (x_2 - 4)^2 + R(5 - x_1 - x_2)^2 u(-g(\boldsymbol{x}))$$

- Thus, when $g(\boldsymbol{x}) < 0$, *i.e.* the decision variables are infeasible, then a penalty of $R(5 - x_1 - x_2)^2$ is applied.

- Proceeding analytically to find the necessary conditions for a minimum, we have

$$\frac{\partial P}{\partial x_1} = 2(x_1^* - 4) + (2R)(5 - x_1^* - x_2^*)(-1) = 0$$

$$\frac{\partial P}{\partial x_2} = 2(x_2^* - 4) + (2R)(5 - x_1^* - x_2^*)(-1) = 0$$

subtracting these two equations

$$2(x_1^* - 4) - 2(x_2^* - 4) = 0 \Rightarrow x_1^* = x_2^*.$$

From the first above, we get

$$(x_1 - 4) - R(5 - 2x_1) = 0$$

and therefore

$$x_1 = \frac{5R+4}{2R+1}$$

- Increasing the penalty parameter to ∞, we have

$$\lim_{R \to \infty} x_1 = \frac{5}{2}$$

and the constrained minimum is:

$$\boldsymbol{x}^* = \left(\frac{5}{2}, \frac{5}{2}\right)$$

Since the constraint, $g(\boldsymbol{x}^*) = 0$ this implies that the constaint is *tight*.

The unconstrained minimum is at $x = (4, 4)$.

- Now solve using the ***inverse penalty***:

$$P(\boldsymbol{x};R) = f(\boldsymbol{x}) + R[g(x)]^{-1}$$

we have

$$P(\boldsymbol{x};R) = (x_1 - 4)^2 + (x_2 - 4)^2 + R(5 - x_1 - x_2)^{-1}$$

Whether or not $g(\boldsymbol{x}) < 0$, *i.e.* whether or not the decision variables are infeasible, a penalty of $R(5 - x_1 - x_2)^{-1}$ is applied.

- We must make sure that we remain feasible during the execution of any algorithm we may employ.

- Proceeding analytically to find the necessary conditions for a minimum, we have

$$\frac{\partial P}{\partial x_1} = 2(x_1^* - 4) + \frac{R}{(5 - x_1^* - x_2^*)^2} = 0$$

$$\frac{\partial P}{\partial x_2} = 2(x_2^* - 4) + \frac{R}{(5 - x_1^* - x_2^*)^2} = 0$$

subtracting these two equations, we again get $x_1^* = x_2^*$ and we also have

$$4(x_1^*)^3 - 36(x_1^*)^2 + 105x_1^* - 100 + \frac{R}{2} = 0$$

- This equation can be solved, for its roots, and the minimum of $P(\boldsymbol{x};R)$ for particular values of R can be found.

Minimum for different values of *R*

R	x_1^*, x_2^*	$f(\boldsymbol{x}^*)$
100	0.5864	23.3053
10	1.7540	10.0890
1	2.2340	6.32375
0.1	2.4113	5.0479
0.01	2.4714	4.6732
0.001	2.4909	4.5548
0	2.5000	4.5000

- For each value of the penalty parameter, an unconstrained optimization problem must be solved.