# Penalty methods

Lecture 30: Presented by Jacomine Grobler

Stellenbosch Unit for Operations Research in Engineering
Department of Industrial Engineering

# Penalty methods

Penalty methods involve solving a sequence of unconstrained optimisation problems whose solutions are generally infeasible in the context of the original constrained problem.

A penalty is, however, incurred for each violated constraint in the solution to the latter problem. As later and later problems in the sequence of unconstrained problems are solved, these penalties are increased, forcing the sequence of solutions towards the feasible region of the original, constrained optimisation problem.

Since solutions to the sequence of unconstrained problems are not required to be feasible in the context of the original problem, a significant advantage of penalty methods is therefore that they are suitable to solve constrained optimisation problems in which some of the constraints are equalities.

# Penalty methods

A *penalty* for an optimisation problem is any function $\psi : \mathbb{R}^n \mapsto \mathbb{R}$ satisfying

$$\psi(\boldsymbol{x}) \begin{cases} = 0 & \text{if } \boldsymbol{x} \text{ is feasible} \\ > 0 & \text{otherwise.} \end{cases}$$

We may use the *quadratic loss function* for equality constraints

$$\psi(\boldsymbol{x}) = \tfrac{\rho}{2} \sum_{j=1}^{q} |a_j - g_j(\boldsymbol{x})|^2$$

and for inequality constraints

$$\psi(\boldsymbol{x}) = \tfrac{\rho}{2} \sum_{j=1}^{p} [\max\{a_j - g_j(\boldsymbol{x}), 0\}]^2$$

for penalty parameter $\rho$.

# Penalty methods

The quadratic-loss penalty function takes the form

$$\gamma(\boldsymbol{x};p) = -f(\boldsymbol{x}) + \frac{\rho}{2}\sum_{j=1}^{p}[\max\{a_j - g_j(\boldsymbol{x}), 0\}]^2 + \frac{\rho}{2}\sum_{k=1}^{q}[b_k - h_k(\boldsymbol{x})]^2.$$

# Penalty methods

**Algorithm 6.2 (Penalty Method)**

**Input**: An instance $\mathcal{I}$ of (6.1)–(6.4), an initial estimate $\boldsymbol{x}^0$ as to an optimal solution to $\mathcal{I}$, an initial penalty parameter value $\rho^0$, a penalty inflation factor $\delta \in (1, \infty)$, and a search termination criterion

**Output**: An approximation of an optimal solution to $\mathcal{I}$ satisfying the search termination criterion

**Step 1** $k \leftarrow 0$

**Step 2** If the search termination criterion is met, then output $\boldsymbol{x}^k$ and stop

**Step 3** Taking $\boldsymbol{x}^k$ as starting point, minimise the penalty function in (6.27) for the fixed parameter value $\rho = \rho^k$ by invoking any numerical method of Chapter 3

**Step 4** $\mu^{k+1} = \delta\mu^k$, $k \leftarrow k + 1$

**Step 5** Return to Step 2

# Example 6.5 (pp. 192–193)

Consider the problem of

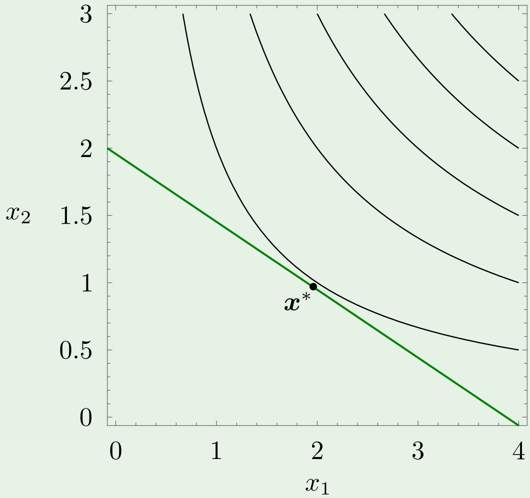
$$\text{maximising} \quad f_{6.4}(x_1, x_2) = x_1 x_2$$

subject to the constraint

$$h_{6.1}(x_1, x_2) = x_1 + 2x_2 = 4.$$

Let us apply Newton's method (Algorithm 3.5) with the initial estimate $\boldsymbol{x}^0 = [0 \ \ 0]^T$. We invoke Algorithm 6.2 with the initial penalty parameter value $\rho^0 = 1$ and the penalty inflation factor $\delta = 10$, terminating the algorithm when no solution component is changing in the first three decimal places any longer (*i.e.* when $||\boldsymbol{p}^k||_1$ drops below $0.000\,5$ for the first time).

# Example 6.5 (pp. 192–193)



# Example 6.5 (pp. 192–193)

| $k$ | $\rho^k$ | $\boldsymbol{x}^k$ | $\nabla\gamma_{6.1}(\boldsymbol{x}^k)$ | $\nabla^2\gamma_{6.1}(\boldsymbol{x}^k)$ | $\boldsymbol{p}^k$ | $\lVert\boldsymbol{p}^k\rVert_1$ |
|---|---|---|---|---|---|---|
| 0 | 1 | $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ | $\begin{bmatrix} -4 \\ -8 \end{bmatrix}$ | $\begin{bmatrix} 1 & 1 \\ 1 & 4 \end{bmatrix}$ | $\begin{bmatrix} 2.666\,667 \\ 1.333\,333 \end{bmatrix}$ | 2.666 667 |
| 1 | 10 | $\begin{bmatrix} 2.666\,667 \\ 1.333\,333 \end{bmatrix}$ | $\begin{bmatrix} 12 \\ 24 \end{bmatrix}$ | $\begin{bmatrix} 10 & 19 \\ 19 & 40 \end{bmatrix}$ | $\begin{bmatrix} -0.615\,385 \\ -0.307\,692 \end{bmatrix}$ | 0.615 385 |
| 2 | 100 | $\begin{bmatrix} 2.051\,282 \\ 1.025\,641 \end{bmatrix}$ | $\begin{bmatrix} 9.230\,769 \\ 18.461\,538 \end{bmatrix}$ | $\begin{bmatrix} 100 & 199 \\ 199 & 400 \end{bmatrix}$ | $\begin{bmatrix} -0.046\,270 \\ -0.023\,135 \end{bmatrix}$ | 0.046 270 |
| 3 | 1 000 | $\begin{bmatrix} 2.005\,013 \\ 1.002\,506 \end{bmatrix}$ | $\begin{bmatrix} 9.022\,556 \\ 18.045\,113 \end{bmatrix}$ | $\begin{bmatrix} 1\,000 & 1\,999 \\ 1\,999 & 4\,000 \end{bmatrix}$ | $\begin{bmatrix} -0.004\,512 \\ -0.002\,256 \end{bmatrix}$ | 0.004 512 |
| 4 | 10 000 | $\begin{bmatrix} 2.000\,500 \\ 1.000\,250 \end{bmatrix}$ | $\begin{bmatrix} 9.002\,251 \\ 18.004\,501 \end{bmatrix}$ | $\begin{bmatrix} 10\,000 & 10\,999 \\ 10\,999 & 40\,000 \end{bmatrix}$ | $\begin{bmatrix} -0.000\,450 \\ -0.000\,225 \end{bmatrix}$ | 0.000 450 |
| 5 | 100 000 | $\begin{bmatrix} 2.000\,050 \\ 1.000\,025 \end{bmatrix}$ | Stop | | | |

# Estimating Lagrange multipliers

It is also possible to obtain estimates of the Lagrange multipliers associated with an optimal solution to the original, constrained optimisation problem.

We may estimate the Lagrange multipliers as

$$\lambda_k(\rho) = \rho\left[h_k(\hat{\boldsymbol{x}}(\rho)) - b_k\right].$$

Let us revisit Example 6.5 to determine the Lagrange multipliers.

# Example 6.6 (pp. 194–195)

Consider the problem of

$$\text{maximising } f_{6.5}\left(x_1, x_2\right) = 2x_1 + 4x_2 - x_1^2 - x_2^2 - 5$$

subject to the constraints

$$x_1 + x_2 = 0$$
$$x_1 - x_2 \geq 1$$