5.7 Penalty method and augmented Lagrangian method

Generic NLP problem

\[
\begin{align*}
\min & \quad f(x) \\
\text{s.t.} & \quad c_i(x) \geq 0 \quad i \in I \\
& \quad c_i(x) = 0 \quad i \in E \\
& \quad x \in \mathbb{R}^n
\end{align*}
\]  \hspace{1cm} (1)

where \( f \) and \( c_i \)'s are of class \( C^1 \) or \( C^2 \).

5.7.1 Quadratic penalty method

**Idea:** Delete constraints, add terms to objective function which penalize the constraint violation, and solve a sequence of resulting unconstrained optimization problems.

Method described for

\[
\begin{align*}
\min & \quad f(x) \\
\text{s.t.} & \quad c_i(x) = 0 \quad i \in E = \{1, \ldots, m\} \\
& \quad x \in \mathbb{R}^n.
\end{align*}
\]  

*(2)*

**Definition:** The **quadratic penalty function** problem associated to (2) is

\[
\begin{align*}
\min_{x \in \mathbb{R}^n} & \quad Q(x, \mu) = f(x) + \frac{1}{2\mu} \sum_{i \in E} c_i^2(x) \\
\end{align*}
\]  

*(3)*

where \(\mu > 0\) is the penalty parameter.

We consider \(\{\mu_k\}_{k \geq 1}\) with \(\lim_{k \to \infty} \mu_k = 0\) and, for each \(k\), we determine an approximate solution \(x_k\) of (3) by using an unconstrained optimization method.
Example:

$$\begin{align*}
\min & \quad x_1 + x_2 \\
\text{s.t.} & \quad x_1^2 + x_2^2 - 2 = 0
\end{align*}$$

with optimal solution \((-1, -1)^t\).

Quadratic penalty problem:

$$
\min_{x \in \mathbb{R}^2} Q(x, \mu) = x_1 + x_2 + \frac{1}{2\mu} (x_1^2 + x_2^2 - 2)^2.
$$

For $\mu = 1$ the minimizer of $Q(x, 1)$ is close to \((-1.1, -1.1)^t\).

For $\mu = 0.1$ the minimizer of $Q(x, 0.1)$ is much closer to \((-1, -1)^t\).

Figures with contours of $Q(x, \mu)$ for different values of $\mu$:

General scheme

0) Select $\varepsilon > 0, \mu_0 > 0$, sequence of tolerances $\{\tau_k\}_{k \geq 0}$ with $\tau_k > 0$ and $\lim_{k \to \infty} \tau_k = 0$. Choose initial $x^s_0$ and set $k = 0$.

1) Determine an approximate minimizer $x_k$ of $Q(x, \mu_k)$ starting from $x^s_k$ and terminate when $||\nabla Q(x, \mu_k)|| \leq \tau_k$.

2) If termination condition is satisfied (e.g, $|f(x_{k-1}) - f(x_k)| < \varepsilon$)

   Then return solution $x_k$

   Else choose $\mu_{k+1} \in (0, \mu_k)$ and starting $x^s_{k+1}$, set $k = k + 1$ and Goto 1)

Choices:

- For convergence results, it suffices that $\lim_{k \to \infty} \tau_k = 0$.

- $\{\mu_k\}_{k \geq 0}$ generated adaptively starting from $\mu_0$: if minimization of $Q(x, \mu_k)$ is "difficult" set e.g. $\mu_{k+1} = 0.7\mu_k$, otherwise $\mu_{k+1} = 0.1\mu_k$.

- Judicious choice of the starting $x^s_k$ when solving unconstrained penalty problem at each iteration: $x^s_{k+1} := x_k$. 
Convergence

**Theorem 1**: Suppose each $x_k$ is a *global minimizer* of $Q(x, \mu_k)$ and $\lim_{k \to \infty} \mu_k = 0$, then every limit point $\bar{x}^*$ of $\{x_k\}_{k \geq 0}$ generated according to the above scheme (with $\tau_k = 0$, $\forall k \geq 0$) is a *global minimum* of problem (2).

**Proof:**
Since in general unconstrained penalty problems are solved approximately, the following is more relevant in practice.

**Theorem 2:** If

- tolerances $\tau_k > 0$ satisfy $\lim_{k \to \infty} \tau_k = 0$
- penalty parameters satisfy $\lim_{k \to \infty} \mu_k = 0$,

then every limit point $\bar{x}^*$ of $\{x_k\}_{k \geq 0}$ at which all $\nabla c_i(\bar{x}^*)$, with $i \in E$, are linearly independent is a *KKT point* of problem (2).

For such points, the subsequence defined by $\mathcal{K}$ with $\lim_{k \in \mathcal{K}} x_k = \bar{x}^*$ satisfies

$$\lim_{k \in \mathcal{K}} -\frac{c_i(x_k)}{\mu_k} = u_i^* \quad \forall i \in E,$$

(4)

where $u^*$ satisfies with $\bar{x}^*$ the KKT conditions for problem (2).
Observation: (4) implies that

i) The minimizer $x_k$ of $Q(x, \mu_k)$ does not satisfy $c_i(x) = 0$ exactly, for all $i \in E$, namely $c_i(x_k) = -\mu_k u^*_i$ for all $i \in E$. To obtain a feasible solution, $\mu_k$ must be driven to 0.

ii) In some circumstances $-\frac{c_i(x_k)}{\mu_k}$ may be used as estimates of Lagrange multipliers $u^*_i$.

Recall: Lagrange function associated to the problem (2) is

$$L(x, u) = f(x) - \sum_{i=1}^{m} u_i c_i(x)$$

and KKT conditions require that, apart from $c_i(x) = 0$ for every $i \in E$,

$$\nabla_x L(x, u) = \nabla f(x) - \sum_{i=1}^{m} u_i \nabla c_i(x) = 0.$$  \hspace{1cm} (6)

By comparing

$$\nabla_x Q(x, \mu) = \nabla f(x) + \frac{1}{\mu} \sum_{i=1}^{m} c_i(x) \nabla c_i(x) = 0$$

and (6), it appears that $-\frac{c_i(x)}{\mu}$ has been substituted with $u_i$. 
It can be proved that if $\tau_k \to 0$ then $x_k \to x^*$ and $-\frac{c_i(x_k)}{\mu_k} \to u_i^*$, $i = 1, 2, \ldots, m$.

**Observation:** When $\mu_k \to 0$ the quadratic penalty problem (3) becomes ill conditioned.

\[
\nabla_{xx}^2 Q(x, \mu_k) = \nabla^2 f(x) + \frac{1}{\mu_k} A^t(x) A(x) + \frac{1}{\mu_k} \sum_{i=1}^{m} c_i(x) \nabla^2 c_i(x) \tag{8}
\]

where $A^t(x) = [\nabla c_1(x), \ldots, \nabla c_m(x)]$ and $A$ is $m \times n$ and of full rank $m \leq n$, with usually $m < n$.

When $x$ is close to minimizer of $Q(x, \mu_k)$ and assumptions of Theorem 2 are satisfied, (4) implies that

\[
\nabla_{xx}^2 Q(x, \mu_k) \approx \nabla_{xx}^2 L(x, u^*) + \frac{1}{\mu_k} A^t(x) A(x). \tag{9}
\]

Since $\nabla_{xx}^2 L(x, u^*)$ does not depend on $\mu_k$ and $\frac{1}{\mu_k} A^t(x) A(x)$ has $n - m$ eigenvalues of value 0 and $m$ eigenvalues of value $O(1/\mu_k)$, numerical issues arise when $\mu_k \to 0$. 
Problems with both equality and inequality constraints:

Quadratic penalty problem

\[
\min_{x \in \mathbb{R}^n} Q(x, \mu) = f(x) + \frac{1}{2\mu} \sum_{i \in E} c_i^2(x) + \frac{1}{2\mu} \sum_{i \in I} ([c_i(x)]^-)^2
\]

(10)

where \([y]^−\) denotes \(\max(-y, 0)\).

Other penalty functions are also available.

If only equality constraints \(c_i(x) = 0, i \in E\), the exact penalty problem is

\[
\min_{x \in \mathbb{R}^n} Q(x, \mu) = f(x) + \frac{1}{2\mu} \sum_{i \in E} |c_i(x)|.
\]

(11)

N.B.: \(Q\) is not everywhere differentiable.