Penalty, Logarithmic barrier methods

- Penalty method
- Logarithmic barrier method

**Goal:** add to the original objective function an extra term that is *zero* when constraints hold and *positive* when constraints do not hold.

**Quadratic penalty**

\[
\phi = f(x) + \frac{1}{2\mu} \sum_{i \in \mathcal{E}} c_i^2
\]

with inequalities

\[
\phi = f(x) + \frac{1}{2\mu} \sum_{i \in \mathcal{E}} c_i^2 + \frac{1}{2\mu} \sum_{i \in \mathcal{E}} (|c_i|)^2, \quad [y^+] = \max(0, y)
\]

**Exact penalty**

\[
\phi = f(x) + \frac{1}{\mu} \sum_{i \in \mathcal{E}} |c_i|
\]

**Example**

\[
\min x_1 + x_2 \text{ subject to } x_1^2 + x_2^2 - 2 = 0, \quad i \in \mathcal{E}
\]

\[
\phi = x_1 + x_2 + \frac{1}{2\mu} (x_1^2 + x_2^2 - 2)^2
\]

not exact but as \(\mu \to 0\) converges to the solution. KKT

\[
1 + \frac{1}{\mu} x_1 (x_1^2 + x_2^2 - 2) = 0,
\]

\[
1 + \frac{1}{\mu} x_2 (x_1^2 + x_2^2 - 2) = 0.
\]

Hence \(x_1 = x_2\) and we solve

\[
1 + \frac{2}{\mu} x_1 (x_1^2 - 1) = 0.
\]

An *exact* augmented Lagrangian

\[
x_1 + x_2 + \lambda (x_1^2 + x_2^2 - 2) + \frac{1}{2\mu} (x_1^2 + x_2^2 - 2)^2
\]

KKT

\[
1 + 2\lambda x_1 + \frac{1}{\mu} x_1 (x_1^2 + x_2^2 - 2) = 0,
\]

\[
1 + 2\lambda x_2 + \frac{1}{\mu} x_2 (x_1^2 + x_2^2 - 2) = 0.
\]
Theorem Suppose, as \( \mu \to 0 \) exact minimizers \( x^*_k \) of the quadratic penalty function \( \phi(x, \mu) \) converge a point \( x \). Then \( x = x^* \) of the original problem.

Note There may be a situation when, say \( x_{2k} \to x_1^* \) and \( x_{2k+1} \to x_2^* \).

Proof Suppose \( \bar{x} \) is a global solution. Then for any \( x_k^* \)

\[
f(x_k^*) + \frac{1}{2\mu} \sum c_i^2(x_k^*) \leq f(\bar{x}) + \frac{1}{2\mu} \sum c_i^2(\bar{x})
\]

Why?

Since \( x_k^* \to x \)

\[
\sum c_i^2(x) = \lim \sum c_i^2(x_k^*) \leq \lim 2\mu_k[f(\bar{x}) - f(x_k)]
\]

Note that \( x \neq \bar{x} \)

An algorithmic framework is choose a new \( \mu_k \to 0 \) as one resolves the minimum of the quadratic function with tolerance

\[
||\nabla \phi(x_k, \mu_k)|| \leq \tau_k, \quad \tau_k \to 0.
\]

Theorem If \( x_k \) satisfy the algorithmic framework, then all limit points \( x_k \) (subsequences) satisfy the KKT conditions of the original problem, moreover

\[
\lim_{k_n \to \infty} -c_i(x_k) \mu_k = \lambda_i^*, \quad \text{for all } i \in \mathcal{E}.
\]

if the \( \nabla c_i(x^*) \) are linearly independent.

Proof

\[
 ||\nabla f(x_k) + \sum \frac{c_i(x_k)}{\mu_k} \nabla c_i(x_k)|| \leq \tau_k \\
 ||\nabla c_i(x_k) \nabla c_i(x_k)|| \leq \mu_k [\tau_k + ||\nabla f(x_k)||] \to 0, \quad \text{since } \mu_k \to 0.
\]

By linear independence \( c_i(x^*) = 0 \) If \( \lambda_i^k = -c_i(x_k)/\mu_k \), it solves

\[
A(x_k)^t \lambda^k = \nabla f(x_k) - \nabla_x \phi(x_k, \mu_k)
\]

or

\[
\lambda^k = [A(x_k)A(x_k)^t]^{-1}A(x_k)[\nabla f(x_k) - \nabla_x \phi(x_k, \mu_k)]
\]

Hence the limit \( \lambda_i^k = -c_i(x_k)/\mu_k \to \lambda^* \) exists:

\[
\lambda^* = [A(x^*)A(x^*)^t]^{-1}A(x^*)\nabla f(x^*).
\]

Note the ill-conditioning of the Hessian of the quadratic penalty function.

Near the iteration point

\[
\nabla_{xx}^2 \phi(x, \mu_k) \approx \nabla_{xx}^2 \mathcal{L}(x, \lambda^*) + \frac{1}{\mu_k} A(x)^t A(x)
\]

We can deal with this ill-conditioning by introducing a dummy vector \( \zeta \)

\[
A(x)p = \mu_k \zeta,
\]

then

\[
\begin{pmatrix} \nabla^2 f + \sum \frac{c_i}{\mu_k} \nabla^2 c_i & A^t \Delta x \\ -\mu_k & \zeta \end{pmatrix} = \begin{pmatrix} -\nabla_x \phi \\ 0 \end{pmatrix}
\]
Barrier function with the barrier parameter

\[ \phi = f(x) - \mu \sum_{i \in I} \log c_i \]

Example

\[ \min x \text{ subject to } x \geq 0, 1 - x \geq 0 \]

\[ \phi = x - \mu \log x - \mu \log(1 - x). \]

An algorithmic framework is choose a new \( \mu_k \to 0 \) as one resolves the minimum of the quadratic function with tolerance

\[ ||\nabla \phi(x_k, \mu_k)|| \leq \tau_k, \quad \tau_k \to 0. \]

Informal Theorem Suppose \( x^* \) is a local solution of the full problem with KKT, LICQ, strict complimentarity (only one of \( \lambda_i^* \) or \( c_i(x^*) \) is zero) and second order sufficient condition satisfied. Then there is a convergence.

Numerical issues

Poor scaling the log-barrier case is present

\[
\begin{align*}
\nabla_x \phi &= \nabla f - \sum \frac{\mu}{c_i} \nabla c_i \\
\nabla^2_{xx} \phi &= \nabla^2 f - \sum \frac{\mu}{c_i^2} \nabla^2 c_i + \sum \frac{\mu}{c_i} \nabla c_i [\nabla c_i]^t \\
\lambda_i^* &\approx \frac{\mu}{c_i(x)}, \quad \nabla^2_{xx} \phi = \nabla^2 f + \sum \frac{\mu}{c_i^2} \nabla c_i [\nabla c_i]^t
\end{align*}
\]