Lecture Outline

- Quadratic penalty method
- Nonsmooth exact penalty methods
- Method of multipliers or augmented Lagrangian method

You should be able to . . .

- Understand the reasoning behind using penalty methods;
- Identify different penalty methods;
- Cast a problem using penalty methods;
- Implement different penalty methods and characterise their convergence properties.

Penalty and Augmented Lagrangian Methods

- The original problem replaced by a sequence of subproblems where the constraints are represented by terms added to the objective.
  - The *quadratic penalty method* adds a multiple of the square of the violation of each constraint to the objective.
  - In *nonsmooth exact penalty methods*, a single unconstrained problem (rather than a sequence) takes the place of the original constrained problem.
  - The *method of multipliers* or *augmented Lagrangian method* Lagrange multiplier estimates are used to avoid the ill-conditioning that is inherent in the quadratic penalty function.
- A related method is the *log-barrier method* that will be discussed in the interior point methods.

The Quadratic Penalty Method

- The original constrained optimisation problem is replaced by a single function consisting of
  - the original objective of the constrained optimization problem, plus
  - one additional term for each constraint, which is positive when the current point x violates that constraint and zero otherwise.
- Often a sequence of such penalty functions are used.
- Consider $\min_x \quad f(x), \quad \text{s.t.} \quad c_i(x) = 0, i \in \mathcal{E}$.
- The quadratic penalty function $Q(x;\mu)$ for this problem is

$$Q(x;\mu) = f(x) + \frac{\mu}{2} \sum_{i \in \mathcal{E}} c_i^2(x),$$

- $\mu > 0$ is the penalty parameter. By driving μ to ∞ the constraint violations become more severe.
- Let $\{\mu_k\}$ where $\mu_k \uparrow \infty$ as $k \to \infty$. At each step the approximate minimiser x_k of $Q(x;\mu_k)$ is found.

The Quadratic Penalty Method

Algorithm: Quadratic Penalty Method

Choose x_0, $k \leftarrow 0$, and sequences $\{\mu_k\} \uparrow \infty$ and $\{\tau_k\} \to 0$
while A termination condition is not satisfied **do**
  $x_{k+1} \leftarrow \arg\min_x Q(x; \mu_k)$ ▷ The termination condition for the subproblem: $\|\nabla_x Q(x; \mu_k)\| \leq \tau_k$.
  $k \leftarrow k + 1$
end while

- $\{\mu_k\}$ can be chosen adaptively based on the difficulty of minimizing the penalty function at each iteration.
- $\|\nabla_x Q(x; \mu_k)\| \leq \tau_k$ may not be satisfied because the iterates may move away from the feasible region when the penalty parameter is not large enough.
- A practical implementation increases the penalty parameter or restores the initial point when
  - the constraint violation is not decreasing rapidly enough, or
  - when the iterates appear to be diverging.

The Quadratic Penalty Method

- As μ_k becomes large the Hessian $\nabla^2 Q(x;\mu_k)$ becomes ill-conditioned near the minimizer.
- This results in the poor performance of many algorithms, e.g. quasi-Newton algorithms
- Newton's method itself is not sensitive to ill conditioning of the Hessian, but
  - The ill conditioning of $\nabla^2 Q(x;\mu_k)$ causes numerical problems when solving for the Newton step (There is a way around it).
  - Even when x is close to $x^\star$, the minimizer of $Q(x;\mu_k)$, the quadratic Taylor series approximation to $Q(x;\mu_k)$ about x is a reasonable approximation of the true function only in a small neighborhood of x.
- The second issue can be mitigated by decreasing the grow rate of μ_k.

The Quadratic Penalty Method

Theorem: *Suppose $Q(x;\mu_k)$ has a (finite) minimiser for each value of μ_k and each x_k is its exact global minimiser, and that $\{\mu_k\} \uparrow \infty$. Then every limit point $x^\star$ of the sequence $\{x_k\}$ is a global solution of the original problem.*

- Let $\bar{x}$ be the global minimiser of the original problem. Since x_k is its exact global minimiser of $Q(x;\mu_k)$:

$$Q(x_k;\mu_k) \le Q(\bar{x};\mu_k) \Rightarrow f(x_k)+\frac{\mu}{2}\sum_{i\in\mathcal{E}} c_i^2(x_k) \le f(\bar{x})+\frac{\mu}{2}\sum_{i\in\mathcal{E}} c_i^2(\bar{x})$$

- Since, $c_i^2(\bar{x}) = 0$: $\sum_{i\in\mathcal{E}} c_i^2(x_k) \le \frac{2}{\mu}(f(\bar{x}) - f(x_k))$.
- There is infinite subsequence $\mathcal{K}$: $\lim_{k\in\mathcal{K}} x_k = x^\star$. Thus,

$$\sum_{i\in\mathcal{E}} c_i^2(x^\star) = \lim_{k\in\mathcal{K}} \sum_{i\in\mathcal{E}} c_i^2(x_k) \le \lim_{k\in\mathcal{K}} \frac{2}{\mu}(f(\bar{x}) - f(x_k)) = 0.$$

- Thus $x^\star$ is feasible.

The Quadratic Penalty Method

- Similarly, $f(x^\star) \leq f(x^\star) + \lim_{k \in \mathcal{K}} \dfrac{\mu}{2} \sum_{i \in \mathcal{E}} c_i^2(x_k) \leq f(\bar{x})$
- Hence, $x^\star$ is a global minimiser as well. $\square$
- This result requires finding the global minimiser at each step and it might not be true in general.

Theorem: *Suppose $\{\mu_k\} \uparrow \infty$ and $\{\tau_k\} \to 0$, and that $\|\nabla_x Q(x; \mu_k)\| \leq \tau_k$ is satisfied at each k. Then if a limit point $x^\star$ of $\{x_k\}$ is infeasible, it is a stationary point of the function $\|c(x)\|^2$. On the other hand, if a limit point $x^\star$ is feasible and the constraint gradients $\nabla c_i(x^\star)$ are linearly independent, then $x^\star$ is a KKT point for the original problem. For such points, we have for any infinite subsequence $\mathcal{K}$ such that $\lim_{k \in \mathcal{K}} x_k = x^\star$ that $\lim_{k \in \mathcal{K}} -\mu_k c_i(x_k) = \lambda_i^\star$, $\forall i \in \mathcal{E}$, where $\lambda^\star$ is the multiplier vector that satisfies the KKT conditions for the original problem.*

The Quadratic Penalty Method

- From the definition of the Q and the termination condition:

$$\nabla_x Q(x_k;\mu_k) = \nabla f(x_k) + \mu_k \sum_{i\in\mathcal{E}} c_i(x_k)\nabla c_i(x_k)$$

$$\left\| \nabla f(x_k) + \mu_k \sum_{i\in\mathcal{E}} c_i(x_k)\nabla c_i(x_k) \right\| \leq \tau_k \overset{(\|a\|-\|b\|\leq\|a+b\|)}{\Longrightarrow}$$

$$\left\| \sum_{i\in\mathcal{E}} c_i(x_k)\nabla c_i(x_k) \right\| \leq \frac{1}{\mu_k}(\tau_k + \|\nabla f(x_k)\|)$$

- Given a subsequence $\mathcal{K}$ where $\lim_{k\in\mathcal{K}} x_k = x^\star$, then $\sum_{i\in\mathcal{E}} c_i(x^\star)\nabla c_i(x^\star) = 0$. Thus, $x^\star$ is a stationary point of $\|c_i(x)\|^2$. If $\nabla c_i(x^\star)$ are independent then $x^\star$ is feasible.

The Quadratic Penalty Method

- Let $C(x)^T = [\nabla c_i(x)]_{i \in \mathcal{E}}$ and $\lambda_k = -\mu_k c(x_k)$:

$$C(x_k)^T \lambda_k = \nabla f(x_k) - \nabla Q(x_k; \mu_k), \quad \|\nabla_x Q(x; \mu_k)\| \leq \tau_k$$

- For all sufficiently large $k \in \mathcal{K}$, $C(x_k)$ is full row rank:

$$\lambda_k = \left[C(x_k) C(x_k)^T\right]^{-1} C(x_k)(\nabla f(x_k) - \nabla Q(x_k; \mu_k))$$

- Hence, $\lim_{k \in \mathcal{K}} \lambda_k = \lambda^\star = \left[C(x^\star) C(x^\star)^T\right]^{-1} C(x^\star) \nabla f(x^\star)$.
- Consequently, by taking limits of the termination condition:

$$\nabla f(x^\star) - C(x^\star)^T \lambda^\star = 0$$

- So, the KKT conditions are satisfied. □
- If $x^\star$ is not feasible, it is at least a stationary point for the function $\|c(x)\|^2$.

The Quadratic Penalty Method: Ill Conditioning and Reformulation

- We examine the nature of the ill conditioning in the Hessian

$$\nabla^2 Q(x;\mu_k) = \nabla^2 f(x_k) + \mu_k \sum_{i\in\mathcal{E}} c_i(x)\nabla^2 c_i(x) + \mu_k C(x)^T C(x)$$

- When x is close to the minimiser of $\nabla^2 Q(x;\mu_k)$ from the previous theorem:

$$\nabla^2 Q(x;\mu_k) \approx \nabla^2 \mathbf{L}(x,\lambda^\star) + \mu_k C(x)^T C(x)$$

- Thus, $\nabla^2 Q(x;\mu_k)$ is equal to the sum of
  - a matrix whose elements are independent of μ_k (the Lagrangian term), and
  - a matrix of rank $|\mathcal{E}|$ (often $|\mathcal{E}| < n$) whose nonzero eigenvalues are of order μ_k (the second term)
- $\kappa(\nabla^2 Q(x;\mu_k)) \to \infty$.

The Quadratic Penalty Method: Ill Conditioning and Reformulation

- This ill conditioning makes the computation of the Newton step p problematic: $\nabla^2 Q(x_k;\mu_k)p = -\nabla Q(x_k;\mu_k)$
- Introduce the auxiliary variable ζ, then

$$\begin{bmatrix} \nabla^2 f(x_k) + \sum_{i\in\mathcal{E}} \mu_k c_i(x_k)\nabla^2 c_i(x_k) & C(x_k)^T \\ C(x_k) & -\frac{1}{\mu_k} I \end{bmatrix} \begin{bmatrix} p \\ \zeta \end{bmatrix} = \begin{bmatrix} -\nabla Q(x_k;\mu_k) \\ 0 \end{bmatrix}$$

- When x is not too far from the solution $x^\star$, the coefficient matrix in this system does not have large singular values (of order μ_k).
- So it is a well-conditioned reformulation.
- Neither system may yield a good search direction p (because $\mu_k c_i(x_k)$ might be poor approximations of $-\lambda_i^\star$)
- The reformulation requires solving a linear system of dimension $n + |\mathcal{E}|$ instead of n.

Nonsmooth Penalty Functions

- Some penalty functions are exact: for certain choices of their penalty parameters, a single minimization can yield the exact solution of the original problem.
- A popular one is the ℓ_1 *penalty function* (let $[x]^- = \max(0, -x)$):

$$\phi_1(x;\mu) = f(x) + \mu \sum_{i \in \mathcal{E}} |c_i(x)| + \mu \sum_{i \in \mathcal{I}} [c_i(x)]^-$$

- The penalty is μ times the ℓ_1 norm of the constraint violation.
- The exactness requirement is loosely as the following:
- At $x^\star$, any move into the infeasible region is penalized sharply enough that it produces an increase in the penalty function to a value greater than $\phi_1(x^\star;\mu) = f(x^\star)$, thereby forcing the minimizer of $\phi_1(x;\mu)$ to be at $x^\star$.

Nonsmooth Penalty Functions

Theorem (Exactness of the ℓ_1 Penalty Function): *Suppose that $x^\star$ is a strict local minimiser at which the first-order necessary KKT conditions are satisfied, with Lagrange multipliers $\lambda_i^\star$, $i \in \mathcal{E} \cup \mathcal{I}$. Then $x^\star$ is a local minimiser of $\phi_1(x;\mu)$ for all $\mu > \mu^\star$, where*

$$\mu^\star = \|\lambda^\star\|_\infty = \max_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i^\star.$$

If in addition, the second-order sufficient conditions hold and $\mu > \mu^\star$, then $x^\star$ is a strict local minimizer of $\phi_1(x;\mu)$.

Definition: *A point $\hat{x}$ is a stationary point for the penalty function $\phi_1(x;\mu)$ if $\mathbf{D}(\phi_1(\hat{x};\mu);p) \geq 0$, $\forall p$. Similarly, $\hat{x}$ is a stationary point of the measure of infeasibility $h(x) = \sum_{i \in \mathcal{E}} |c_i(x)| + \sum_{i \in \mathcal{I}} [c_i(x)]^-$ if $\mathbf{D}(h(\bar{x});p) \geq 0$, $\forall p$.*

Nonsmooth Penalty Functions

Definition: *If a point is infeasible for the penalty function but stationary with respect to the infeasibility measure h, we say that it is an infeasible stationary point.*

- Now it is shown that that stationary points of $\phi_1(x;\mu)$ correspond to KKT points of the original constrained optimization problem under certain assumptions.

Theorem: *Suppose that $\hat{x}$ is a stationary point of the penalty function $\phi_1(x;\mu)$ for all $\mu > \hat{\mu} > 0$ for some $\hat{\mu}$. Then, if $\hat{x}$ is feasible for the original problem, then it satisfies the KKT conditions for the original problem. If $\hat{x}$ is not feasible for the original problem, it is an infeasible stationary point.*

Nonsmooth Penalty Functions

Algorithm: Classical ℓ_1 Penalty Method

Choose x_0, $k \leftarrow 0$, $\mu_0 > 0$ and tolerance $\tau > 0$
while $h(x_0) > \tau$ **do**
  Set x_{k+1} to be an approximate minimizer of $\phi_1(x;\mu)$ ▷ Iterations for the solution start at x_k.
  Choose μ_{k+1} ▷ $\mu_{k+1} > \mu_k$, Simplest: $\mu_{k+1} = \gamma\mu_k$
end while

- The minimization of $\phi_1(x;\mu)$ is made difficult by the nonsmoothness of the function. But possible to use a smooth model of it.
- This scheme sometimes works well in practice, but can also be inefficient.
    - Too many cycles if μ_0 is too small.
    - The iterates may move away from the solution (the subproblem should be terminated early, or x to be reset.)
    - The penalty function will be difficult to minimize if μ_k is too large.

Nonsmooth Penalty Functions: A Practical ℓ_1 Penalty Method

- We won't use the generic nonsmooth tehcniques, e.g. bundle methods, but take advantage the particular nondifferentiabilities of the funciton.
- A step toward the minimizer of $\phi_1(x;\mu)$ is obtained via forming a simplified model of it and minimising it.
- It is done by linearizing the constraints c_i and replacing the nonlinear programming objective f by a quadratic function:

$$
\begin{aligned}
Q(p,\mu) = {} & f(x) + \nabla f(x)^T p + \frac{1}{2} p^T W p \\
& + \mu \sum_{i \in \mathcal{E}} |c_i(x) + \nabla c_i(x)^T p| + \mu \sum_{i \in \mathcal{I}} [c_i(x) + \nabla c_i(x)^T p]^-
\end{aligned}
$$

- W is a symmetric matrix which contains second derivative information about f and c_i.

Nonsmooth Penalty Functions: A Practical ℓ_1 Penalty Method

- $Q(p, \mu)$ is not smooth, but minimising it can be formulated as a smooth problem:

$$\begin{aligned}
\min_{p,r,s,t} \quad & f(x) + \frac{1}{2}p^T W p + \nabla f(x)^T p + \mu \sum_{i\in\mathcal{E}} (r_i + s_i) + \mu \sum_{i\in\mathcal{I}} t_i \\
\text{s.t.} \quad & c_i(x) + \nabla c_i(x)^T p = r_i - s_i,\ i \in \mathcal{E} \\
& c_i(x) + \nabla c_i(x)^T p \geq -t_i,\ i \in \mathcal{I} \\
& r, s, t \geq 0
\end{aligned}$$

- This subproblem can be solved with a standard quadratic programming solver.
- Sometimes the penalty parameter is chosen at every iteration so that $\mu_k > \|\lambda_k\|_\infty$, where λ_k is an estimate of the Lagrange multipliers computed at x_k.
- These methods fell out of favour but there has been a resurgence of interest in penalty methods.

Augmented Lagrangian Method: Equality Constraint

- It is related to the quadratic penalty algorithm.
- Better conditioning by introducing explicit Lagrange multiplier estimates into the objective function: *the augmented Lagrangian function*.
- The augmented Lagrangian function largely preserves smoothness.
- The augmented Lagrangian function:

$$\mathbf{L}_A(x, \lambda; \mu) = f(x) - \sum_{i \in \mathcal{E}} \lambda_i c_i(x) + \frac{\mu}{2} \sum_{i \in \mathcal{E}} c_i^2(x)$$

- It is a combination of the Lagrangian function and the quadratic penalty function.

Augmented Lagrangian Method: Equality Constraint

- At each step k and fixing λ_k, from the necessary optimality condition:

$$\nabla \mathbf{L}_A(x_k, \lambda_k; \mu_k) = \nabla f(x_k) - \sum_{i \in \mathcal{E}} \left([\lambda_k]_i - \mu_k c_i(x_k) \right) \nabla c_i(x_k)$$

- Comparing it with the optimality condition of the original problem, ideally:

$$\lambda_i^\star = [\lambda_k]_i - \mu_k c_i(x_k) \implies c(x_k) = -\frac{1}{\mu_k}(\lambda_i^\star - [\lambda_k]_i)$$

- The infeasibility of x_k decreases as the multiplier estimate gets closer to the real one. Also, the infeasibility is much smaller than $1/\mu_k$, compared to being proportional to $1/\mu_k$.
- Thus, to move λ_k closer to $\lambda^\star$:

$$[\lambda_{k+1}]_i = [\lambda_k]_i - \mu_k c_i(x_k)$$

Augmented Lagrangian Method: Equality Constraint

Algorithm: Augmented Lagrangian Method – Equality Constraint

Choose x_0, λ_0, $\mu_0 > 0$, $k \leftarrow 0$, and tolerance τ_0
while A termination condition is not satisfied **do**
  $x_k \leftarrow \arg\min_x \mathbf{L}_A(x, \lambda_k; \mu_k)$ ▷ The starting point and the termination condition for the subproblem: x_k and $\|\nabla_x \mathbf{L}_A(x, \lambda_k; \mu_k)\| \leq \tau_k$.
  $\lambda_{k+1} \leftarrow \lambda_k - \mu_k c(x_k)$
  Choose the penalty parameter μ_{k+1} ▷ $\mu_{k+1} \geq \mu_k$
  Choose tolerance τ_{k+1}
  $k \leftarrow k + 1$
end while

- The convergence can be assured without increasing μ indefinitely; better conditioning
- τ_k may depend on the infeasibility $\|c(x_k)\|_1$.
- μ_k may be increased if the reduction in this infeasibility measure is insufficient.

Augmented Lagrangian Method: Equality Constraint

Theorem: *Let $x^\star$ be a local minimiser at which the LICQ is satisfied, and the second-order sufficient conditions are satisfied for $\lambda = \lambda^\star$. Then there is a threshold value $\bar{\mu}$ such that for all $\mu \geq \bar{\mu}$, $x^\star$ is a strict local minimiser of $\mathbf{L}_A(x, \lambda^\star; \mu)$.*

- The proof is obtained by showing that $x^\star$ satisfies the second-order sufficient condition for $\mathbf{L}_A(x, \lambda^\star; \mu)$ for sufficiently large μ.
- We know that $\mathbf{L}(x^\star, \lambda^\star) = 0$ and $c_i(x^\star) = 0$, then:

$$
\begin{aligned}
\nabla \mathbf{L}_A(x^\star, \lambda^\star; \mu) &= \nabla f(x^\star) - \sum_{i \in \mathcal{E}} \left(\lambda_i^\star - \mu c_i(x^\star)\right) \nabla c_i(x^\star) \\
&= \nabla f(x^\star) - \sum_{i \in \mathcal{E}} \lambda_i \nabla c_i(x^\star) = \mathbf{L}(x^\star, \lambda^\star) = 0.
\end{aligned}
$$

Augmented Lagrangian Method: Equality Constraint

- Let $C(x)^T = [\nabla c_i(x)]_{i\in\mathcal{E}}$:

$$\nabla^2 \mathbf{L}_A(x^\star, \lambda^\star; \mu) = \nabla^2 \mathbf{L}(x^\star, \lambda^\star) + \mu C^T(x^\star) C(x^\star)$$

- To obtain a contradiction assume $\nabla^2 \mathbf{L}_A(x^\star, \lambda^\star; \mu) \not\succ 0$. Then for each $k \geq 1$, $\exists w_k : \|w_k\| = 1$ such that

$$0 \geq w_k^T \nabla^2 \mathbf{L}_A(x^\star, \lambda^\star; k) w_k = w_k^T \nabla^2 \mathbf{L}(x^\star, \lambda^\star) w_k + k\|C(x^\star) w_k\|^2$$

$$\|C(x^\star) w_k\|^2 \leq -\frac{1}{k} w_k^T \nabla^2 \mathbf{L}(x^\star, \lambda^\star) w_k$$

- $\{w_k\}$ are in a compact set and have an accumulation point w. taking the limit $k \to \infty$: $\|C(x_k) w_k\|^2 \to \|C(x^\star) w\|^2$, $C(x^\star) w = 0$. Moreover,

$$w_k^T \nabla^2 \mathbf{L}(x^\star, \lambda^\star) w_k \leq -k\|C(x^\star) w_k\|^2 \leq 0$$

- Taking the limit: $w^T \nabla^2 \mathbf{L}(x^\star, \lambda^\star) w \leq 0$, which contradicts the assumption. □

Augmented Lagrangian Method: Equality Constraint

Theorem: *Suppose that the assumptions of the previous theorem are satisfied at $x^\star$ and $\lambda^\star$ and let $\bar{\mu}$ be chosen as in that theorem. Then there exist positive scalars δ, ϵ and M such that for all λ_k and μ_k satisfying $\|\lambda_k - \lambda^\star\| \leq \mu_k \delta$ for $\mu_k \geq \bar{\mu}$ the following claims hold:*

(a) *The problem*

$$\min_{x} \quad \mathbf{L}_A(x, \lambda_k; \mu_k), \quad s.t. \ \|x - x^\star\| \leq \epsilon,$$

has a unique solution x_k and

$$\|x_k - x^\star\| \leq M\|\lambda_k - \lambda^\star\|/\mu_k.$$

(b) *Let $\lambda_{k+1} = \lambda_k - \mu_k c(x_k)$, then*

$$\|\lambda_k - \lambda^\star\| \leq M\|\lambda_k - \lambda^\star\|/\mu_k.$$

(c) *The constraint gradients $\nabla c_i(x_k)$, $i \in \mathcal{E}$, are linearly independent and $\nabla^2 \mathbf{L}_A(x, \lambda_k; \mu_k) > 0$*

- x_k will be close to $x^\star$ if λ_k is accurate or if μ_k is large.

Practical Augmented Lagrangian Methods: Bound-Constrained Formulation

- Via introducing slack variables, s_i the constraints in $\mathcal{I}$ can be written as $c_i(x) - s_i = 0, \quad s_i \geq 0, \quad \forall i \in \mathcal{I}$.
- Bound constraints $l \leq x \leq u$ can be kept unchanged.
- Then the problem (after incorporating s_i in x and relabeling c_i accordingly) takes the form of

$$\min_x \; f(x), \quad \text{s.t. } c_i(x) = 0, \; i \in \{1, \ldots, m\}, \; l \leq x \leq u$$

- The bound-constrained Lagrangian (BCL):

$$\mathbf{L}_A(x, \lambda; \mu) = f(x) - \sum_{i=1}^{m} \lambda_i c_i(x) + \frac{\mu}{2} \sum_{i=1}^{m} c_i^2(x)$$

- The subproblem captures the bound constraints:

$$\min_x \; \mathbf{L}_A(x, \lambda; \mu), \quad \text{s.t. } l \leq x \leq u$$

- The multipliers λ and the penalty parameter μ are updated and the process is repeated.

Bound-Constrained Formulation

Algorithm: Bound-Constrained Lagrangian Method – `LANCELOT`

Choose x_0, λ_0, and tolerances $\eta^\star$ and $\omega^\star$
$\mu_0 \leftarrow 10$; $\omega_0 \leftarrow 1/\mu_0$; $\eta_0 \leftarrow 1/\mu_0^{0.1}$; $k \leftarrow 0$
while $\|c(x_k)\| > \eta^\star$ and $|x_k - \mathbf{P}(x_k - \nabla \mathbf{L}_A(x_k, \lambda_k; \mu_k), l, u)\| > \omega^\star$ **do**
  Find x_k s.t. $\|x_k - \mathbf{P}(x_k - \nabla \mathbf{L}_A(x_k, \lambda_k; \mu_k), l, u)\| \leq \omega_k$ ▷ $\mathbf{P}(y, l, u)$ projects y into the box defined by lower and upper bounds l and u.
  if $\|c(x_k)\| \leq \eta_k$ **then** ▷ Update multipliers, tighten tolerances.
    $\lambda_{k+1} \leftarrow \lambda_k - \mu_k c(x_k)$
    $\mu_{k+1} \leftarrow \mu_k$; $\eta_{k+1} \leftarrow \eta_k/\mu_{k+1}^{0.9}$; $\omega_{k+1} \leftarrow \omega_k/\mu_{k+1}$;
  else ▷ Increase penalty parameter, tighten tolerances.
    $\lambda_{k+1} \leftarrow \lambda_k$
    $\mu_{k+1} \leftarrow 100\mu_k$; $\eta_{k+1} \leftarrow 1/\mu_{k+1}^{0.1}$; $\omega_{k+1} \leftarrow 1/\mu_{k+1}$;
  end if
end while

Practical Augmented Lagrangian Methods: Linearly Constrained Formulation

- Linearly constrained Lagrangian (LCL) methods is to generate a step by minimizing the Lagrangian subject to linearizations of the constraints.
- The subproblem in LCL:

$$\begin{aligned} \min_x \quad & F_k(x) \\ \text{s.t.} \quad & c(x_k) + C(x_k)(x - x_k) = 0, \quad l \le x \le u \end{aligned}$$

- Choices for $F_k(x)$

$$F_k(x) = f(x) - \sum_{i=1}^{m} [\lambda_k]_i [\bar{c}_k(x)]_i \quad \text{(early)}$$

$$F_k(x) = f(x) - \sum_{i=1}^{m} [\lambda_k]_i [\bar{c}_k(x)]_i + \frac{\mu}{2} \sum_{i=1}^{m} [\bar{c}_k(x)]_i^2 \quad (\text{current} - \texttt{MINOS})$$

$$[\bar{c}_k(x)]_i = c_i(x) - c_i(x_k) - \nabla c_i(x_k)^T (x - x_k)$$

Practical Augmented Lagrangian Methods: Unconstrained Formulation

- Using a derivation based on the proximal point approach.
- For simplicity assume $\mathcal{E} = \emptyset$.
- The the problem is defined as $\min_x \; F(x)$, where

$$F(x) = \max_{\lambda \geq 0} \left(f(x) - \sum_{i \in \mathcal{I}} \lambda_i c_i(x) \right) = \begin{cases} f(x) & x \in \overbrace{\mathcal{X}}^{\text{feasible set}} \\ \infty & \text{otherwise} \end{cases}$$

- F is nonsmooth but can be approximated by a smooth $\hat{F}$:

$$\hat{F}(x; \lambda_k, \mu_k) = \max_{\lambda \geq 0} \left(f(x) - \sum_{i \in \mathcal{I}} \lambda_i c_i(x) - \frac{1}{2\mu_k} \sum_{i \in \mathcal{I}} (\lambda_i - [\lambda_k]_i)^2 \right)$$

- A penalty for any move of λ away from the previous value; λ to stay *proximal* to the previous estimate.

Unconstrained Formulation

$$\hat{F}(x;\lambda_k,\mu_k) = \max_{\lambda\geq 0}\left(f(x) - \sum_{i\in\mathcal{I}} \lambda_i c_i(x) - \frac{1}{2\mu_k}\sum_{i\in\mathcal{I}}(\lambda_i - [\lambda_k]_i)^2 \right)$$

- Maximisaiton is bound-constrained separable quadratic problem in λ_i, thus

$$\lambda_i = \begin{cases} 0 & -c_i(x) + [\lambda_k]_i/\mu_k \leq 0 \\ [\lambda_k]_i - \mu_k c_i(x) & \text{otherwise} \end{cases}$$

- Consequently, $\hat{F}(x;\lambda_k,\mu_k) = f(x) + \sum_{i\in\mathcal{I}} \psi(c_i(x), [\lambda_k]_i; \mu_k)$ where

$$\psi(s,\lambda;\mu) = \begin{cases} -\lambda s + \dfrac{\mu}{2}s^2 & s - \lambda/\mu \leq 0 \\ -\dfrac{1}{2\mu}\lambda^2 & \text{otherwise} \end{cases}$$

- Thus x_k can be obtained by minimising x and λ_{k+1} from the above formula.
- Not implemented in any package yet.