

Numerical Optimisation
Constraint optimisation:

Penalty and augmented Lagrangian methods

Marta M. Betcke

`m.betcke@ucl.ac.uk`

Kiko Rullan

`f.rullan@cs.ucl.ac.uk`

Department of Computer Science,
Centre for Medical Image Computing,
Centre for Inverse Problems
University College London

Lecture 14

Constraint optimization problem

Assignment Project Exam Help

$$\begin{aligned} \min_{x \in D \subseteq \mathbb{R}^n} & f(x) \\ \text{subject to} & f_i(x) \leq 0, \quad i = 1, \dots, m, \\ & h_i(x) = 0, \quad i = 1, \dots, p \end{aligned}$$

<https://powcoder.com>

Conflicting goals: minimise the function and satisfy the constraints.

Idea: Minimise a merit function $Q(x; \mu)$ where μ are parameters. Some minimisers of $Q(x; \mu)$ approach those of f subject to the constraints as μ approach some set \mathcal{M} .

Benefit: reformulation as an unconstraint problem.

Consider a problem with equality constraints

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & f(x) \\ \text{subject to} \quad & h_i(x) = 0, \quad i = 1, \dots, p. \end{aligned} \quad (\text{QPF})$$

The merit function (*quadratic penalty function*)

$$Q(x; \mu) := f(x) + \frac{\mu}{2} \sum_{i=1}^p h_i^2(x),$$

where $\mu > 0$ is the *penalty parameter*.

Idea: choose a sequence $\{\mu_k\} : \mu_k \rightarrow \infty$ as $k \rightarrow \infty$,
i.e. increasingly penalise the constraint, and compute the sequence
 $\{x_k\}$ of (approximate) minimisers of $Q(x; \mu_k)$.

Convergence for the quadratic penalty

Let $\{x_k\}$ be the sequence of approximate minimisers of $Q(x; \mu_k)$, such that $\|\nabla_x Q(x_k; \mu_k)\| \leq \tau_k$, x^* be the limit point of $\{x_k\}$ as the sequences of the penalty parameters $\mu_k \rightarrow \infty$ and tolerances, $\tau_k \rightarrow 0$.

- If a limit point x^* is infeasible, it is a stationary point of $\|h(x)\|^2$.
- If a limit point x^* is feasible and the constraint gradients $\nabla h_i(x^*)$ are linearly independent, then x^* is a KKT point for (COP:E), and we have that

$$\lim_{k \rightarrow \infty} \mu_k h_i(x_k) = \nu_i^*, \quad i = 1, \dots, p,$$

where ν^* is the multiplier vector that satisfies the KKT conditions for (COP:E).

$$\nabla_x Q(x_k; \mu_k) = \nabla f(x_k) + \sum_{i=1}^p \mu_k h_i(x_k) \nabla h_i(x_k) \quad (1)$$

From the convergence criterium $\|\nabla_x Q(x_k; \mu_k)\| \leq \tau_k$ (using the inequality $\|a\| - \|b\| \leq \|a + b\|$) we obtain

$$\left\| \sum_{i=1}^p h_i(x_k) \nabla h_i(x_k) \right\| \leq \frac{1}{\mu_k} (\tau_k + \|\nabla f(x_k)\|).$$

As $k \rightarrow \infty$: $\tau_k \rightarrow 0$, $\|\nabla f(x_k)\| \rightarrow \|\nabla f(x^*)\|$ and $\mu_k \rightarrow \infty$ this

$$\sum_{i=1}^p h_i(x^*) \nabla h_i(x^*) = 0.$$

- i) If $h_i(x^*) \neq 0, i = 1, \dots, p$ then $\nabla h_i(x^*)$ are linearly dependent which implies that x^* is a stationary point of $\|h(x)\|^2$.
- ii) If $\nabla h_i(x^*), i = 1, \dots, p$ are linearly independent, $h_i(x^*) = 0$ and x^* is primarily feasible i.e. satisfies the second KKT condition. It remains to show that the “dual feasibility” (the first KKT condition) is satisfied.

Case ii):

Intuition:

As $k \rightarrow \infty, Q(x^k)$ should approach the Lagrangian

$$\mathcal{L}(x^*; \nu^*) = f(x^*) + \sum_{i=1}^p \nu_i^* h_i(x^*). \quad (2)$$

and $\nabla_x Q(x^k)$ its derivative i.e. the “dual feasibility” condition

$$\nabla_x \mathcal{L}(x^*; \nu^*) = \nabla f(x^*) + \sum_{i=1}^p \nu_i^* \nabla h_i(x^*). \quad (3)$$

Rearranging (1) and denoting $A(x)^T := \nabla h_i(x_k), i = 1, \dots, p$ and $\nu^k := \mu_k h(x_k)$ we obtain

$$A(x_k)^T \nu^k = -\nabla f(x_k) + \nabla_x Q(x_k; \mu_k), \quad \|\nabla_x Q(x_k; \mu_k)\| \leq \tau_k.$$

Assignment Project Exam Help
For large enough k the matrix $A(x_k)$ has full row rank and hence the above overdetermined system has the unique solution

$$\nu^k = (A(x_k)A(x_k)^T)^{-1} A(x_k)[- \nabla f(x_k) + \nabla_x Q(x_k; \mu_k)].$$

Taking the limit as $k \rightarrow \infty$

$$\lim_{k \rightarrow \infty} \nu^k = \nu^* = -(A(x^*)A(x^*)^T)^{-1} A(x^*) \nabla f(x^*)$$

and the same in (1) yields the “dual feasibility” condition

$$\nabla f(x^*) + A(x^*)^T \nu^* = 0.$$

Hence, x^* is the KKT point with unique Lagrange multiplier ν^* .

Example

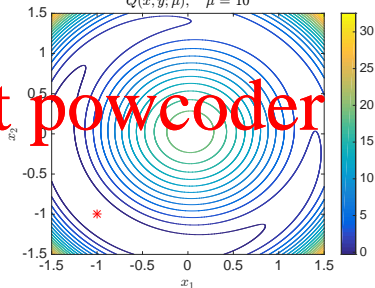
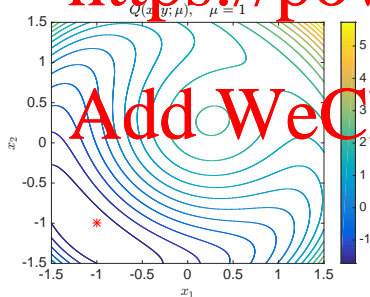
$$\min x_1 + x_2$$

$$\text{subject to } x_1^2 + x_2^2 - 2 = 0.$$

Solution: $(-1, -1)^T$.

Quadratic penalty function: $Q(x; \mu) = x_1 + x_2 + \frac{\mu}{2}(x_1^2 + x_2^2 - 2)^2$.

<https://powcoder.com>



Add WeChat powcoder

Assignment Project Exam Help

$$\begin{aligned} \min \quad & -5x_1^2 + x_2^2 \\ \text{subject to} \quad & x_1 = 1. \end{aligned}$$

Solution: $(1, 0)$. <https://powcoder.com>

Quadratic penalty function: $Q(x; \mu) = -5x_1^2 + x_2^2 + \frac{\mu}{2}(x_1 - 1)^2$.

$Q(x; \mu)$ is unbounded for $\mu < 10$.

Add WeChat powcoder

The iterates would diverge. Unfortunately, a common problem.

Ill-conditioning of Hessian

Newton step: $\nabla_{xx}^2 Q(x; \mu_k) p_n = -\nabla_x Q(x; \mu_k)$

$$\nabla_{xx}^2 Q(x; \mu_k) = \nabla_{xx}^2 f(x) + \sum_{i=1}^p \underbrace{\mu_k h_i(x)}_{\approx \nu_i} \nabla^2 h_i(x) + \mu_k \underbrace{\sum_{i=1}^p h_i(x) \nabla h_i(x)^T}_{=: A(x)^T}$$

If x is sufficiently close to the minimiser of $Q(\cdot; \mu_k)$

$$\nabla_{xx}^2 Q(x; \mu_k) \approx \nabla_{xx}^2 \mathcal{L}(x; \nu) + \mu_k A(x) A(x)^T$$

As $\mu_k \rightarrow \infty$ the Hessian is dominated by the second term and hence increasingly ill-conditioned.

Alternative formulation avoids ill-conditioning, $\zeta = \mu_k A(x) p_n$

$$\begin{bmatrix} \nabla^2 f(x) + \sum_{i=1}^p \mu_k h_i(x) \nabla^2 h_i(x) & A(x)^T \\ A(x) & \mu_k^{-1} I \end{bmatrix} \begin{bmatrix} p_n \\ \zeta \end{bmatrix} = \begin{bmatrix} -\nabla_x Q(x; \mu_k) \\ 0 \end{bmatrix}.$$

Still, if $\mu_k h_i(x)$ is not a good enough approximation to ν^* , inadequate quadratic model yields inadequate search direction p_n .

For general constraint problems including equality and inequality constraints, the quadratic penalty function can be defined as

$$Q(x; \mu) := f(x) + \frac{\mu}{2} \sum_{i=1}^p h_i^2(x) + \frac{\mu}{2} \sum_{i=1}^m ([f_i(x)]^-)^2,$$

where $[y]^- := \max\{y, 0\}$

Note: Q may be less smooth than the objective and constraint functions e.g. $f_1(x) = x_1 \geq 0$, then $\max\{y, 0\}^2$ has discontinuous second derivate and so does Q .

- μ_k can be chosen adaptively based on the difficulty of minimising the penalty function in each iteration i.e. when minimising $Q(x; \mu_k)$ is expensive, choose μ_{k+1} moderately larger than μ_k e.g. $\mu_{k+1} = 1.5\mu_k$, when minimising $Q(x; \mu_k)$ is cheap, choose μ_{k+1} larger e.g. $\mu_{k+1} = 10\mu_k$.

- There is no guarantee that $\|\nabla_x Q(x; \mu_k)\| \leq \tau_k$ will be satisfied. Practical implementations need safe guards to increase μ (and possibly restore the initial point) when constraint violation is not decreasing fast enough or when the iterates appear diverging.

- When only equality constraints are present, $Q(x; \mu_k)$ is smooth and algorithms for unconstrained optimisation can be used, however $Q(x; \mu_k)$ becomes more difficult to minimise as μ_k becomes large unless special techniques are used. In particular, methods like conjugate gradients and quasi-Newton will perform poorly, Newton method can be adapted (see the reformulation of Newton step) but it still can yield inadequate direction due to inadequacy of quadratic approximation.

- Choice of initial point e.g. warm start $x_{k+1}^S = x_k^S$ can improve performance of Newton.

Nonsmooth penalty functions

Some penalty functions are *exact* i.e. for certain choices of penalty parameters, minimisation w.r.t. x yields the exact minimiser of f .

To be exact the function has to be nonsmooth.

Assignment Project Exam Help

$$Q_1(x, \mu) := f(x) + \mu \sum_{i=1}^p |h_i(x)| + \mu \sum_{i=1}^m [f_i(x)]^-,$$

where $[y]^- := \max\{y, 0\}$.

Let x^* be a strict local minimiser of (COP), which satisfies the 1st order necessary conditions with Lagrange multipliers ν^*, λ^* . Then x^* is a local minimiser of $Q_1(x; \mu)$ for all $\mu > \mu^* = \|(\nu^*, \lambda^*)^T\|_\infty$. If moreover the 2nd order sufficient conditions hold at $\mu > \mu^*$, then x^* is a strict local minimiser of $Q_1(x; \mu)$.

Let \hat{x} be a stationary point of the penalty function $Q_1(x; \mu)$ for all $\mu > \hat{\mu} > 0$. Then, if \hat{x} is feasible for (COP), it satisfies KKT conditions. If \hat{x} is not feasible for (COP), it is an infeasible stationary point.

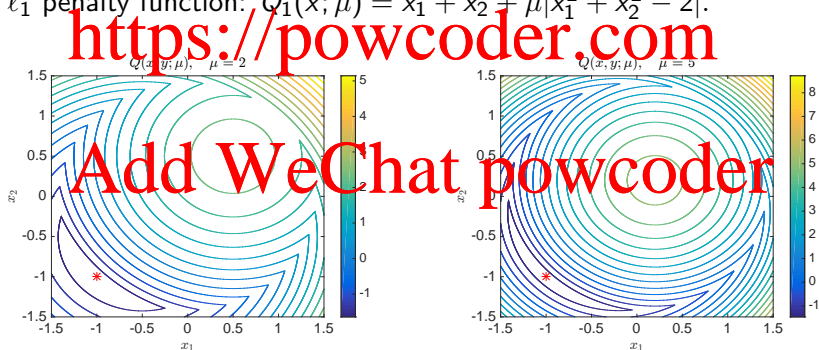
Example revisited

$$\min x_1 + x_2$$

$$\text{subject to } x_1^2 + x_2^2 - 2 = 0.$$

Solution: $(-1, -1)^T$.

ℓ_1 penalty function: $Q_1(x; \mu) = x_1 + x_2 + \mu|x_1^2 + x_2^2 - 2|$.



Augmented Lagrangian

Reduces ill-conditioning by introducing explicit Lagrange multiplier estimates into the function to be minimised.

Can preserve smoothness. Can be implemented using standard unconstrained (or bound constrained) optimization.

Motivation: The minimisers x_k of $Q(x; \mu_k)$ do not quite satisfy the feasibility condition $h_i(x) = 0$

$$h_i(x_k) \approx \nu^*/\mu_k, \quad i = 1, \dots, p.$$

Obviously, in the limit $\mu_k \rightarrow \infty$, $h_i(x) \rightarrow 0$ but can we avoid this systematic perturbation for moderate values of μ_k ?

Augmented Lagrangian:

$$\mathcal{L}_A(x, \nu; \mu) := f(x) + \sum_{i=1}^p \nu h_i(x) + \frac{\mu}{2} \sum_{i=1}^p h_i^2(x).$$

Update of Lagrange multiplier estimate

Optimality condition for the unconstrained minimiser of $\mathcal{L}_A(x, \nu^k; \mu_k)$

$$0 \approx \nabla_x \mathcal{L}_A(x_k, \nu^k; \mu_k) = \nabla f(x_k) + \sum_{i=1}^p [\nu_i^k + \mu_k h_i(x_k)] \nabla h_i(x_k).$$

Optimality condition for the Lagrangian of (COP:E)

$$0 \approx \nabla_x \mathcal{L}(x_k, \nu^*) = \nabla f(x_k) + \sum_{i=1}^p \nu_i^* \nabla h_i(x_k).$$

Comparison yields (an update scheme for ν):

$$\nu^* \approx \nu^k + \mu_k h_i(x_k), \quad i = 1, \dots, p$$

as from $h_i(x_k) = \frac{1}{\mu_k}(\nu_i^* - \nu_i^k)$, $i = 1, \dots, p$ we see that if ν^k is close to ν^* the infeasibility goes to 0 faster than $1/\mu_k$.

Example revisited

$$\begin{aligned} \min \quad & x_1 + x_2 \\ \text{subject to} \quad & x_1^2 + x_2^2 - 2 = 0. \end{aligned}$$

Solution: $(-1, -1)$

Augmented Lagrangian:

$$\mathcal{L}(x, \nu; \mu) = x_1 + x_2 + \nu(x_1^2 + x_2^2 - 2) + \frac{\mu}{2}(x_1^2 + x_2^2 - 2)^2.$$

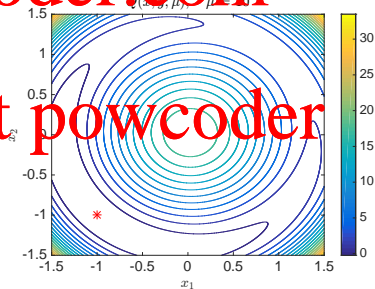
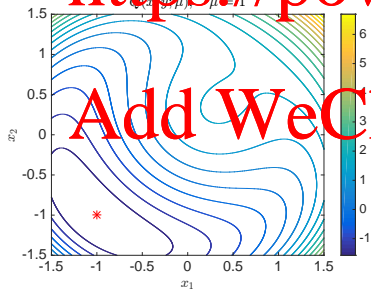


Figure: $\nu = 0.4$

Let x^* be a local minimiser of (COP:E) at which the constraint gradients are linearly independent and which satisfies the 2nd order sufficient conditions with Lagrange multipliers ν^* . Then for all

$\mu \geq \bar{\mu} > 0$, x^* is a strict local minimiser of $L_A(x, \nu^*; \mu)$. Furthermore, there exist $\delta, \epsilon, M > 0$ such that for all ν^k, μ_k satisfying

$$\|\nu^k - \nu^*\| \leq \mu_k \delta, \quad \mu_k \geq \bar{\mu},$$

- the problem $\min L_A(x, \nu^k; \mu_k)$, subject to $\|x - x^*\| \leq \epsilon$, has a unique solution x_k and it holds

- it holds

$$\|x_k - x^*\| \leq M \|\nu^k - \nu^*\| / \mu_k$$
$$\|\nu^{k+1} - \nu^*\| \leq M \|\nu^k - \nu^*\| / \mu_k,$$

where $\nu^{k+1} = \nu^k + \mu_k h(x_k)$.

- the matrix $\nabla_{xx}^2 \mathcal{L}_A(x_k, \nu^k; \mu_k)$ is positive definite and the constraint gradients $\nabla h_i(x_k), i = 1, \dots, p$ are linearly independent.

Assignment Project Exam Help

- **Bound constraint formulation:** convert inequality constraints into equality constraints using slack variables

$$f_i(x) - s_i = 0, \quad s_i \geq 0, \quad i \in \{1, \dots, m\}.$$

<https://powcoder.com>
Bound constraints are not transformed. Solve by projected gradient algorithm

Add WeChat [powcoder](https://powcoder.com)

where $P(\cdot; l, u)$ projects on the box $[l, u]$.

- **Linearly constraint formulation:** transform into equality constraint problem and linearise constraints

$$\min F_k(x), \text{ subject to } f_i(x_k) + \nabla f_i^T(x_k)(x - x_k) = 0, \quad l \leq x \leq u.$$

Choose F_k as

$$F_k(x) = f(x) + \sum_{i=1}^m \nu_i^k \bar{f}_i^k(x),$$

where

$$\bar{f}_i^k(x) = f_i(x) - f_i(x_k) - \nabla f_i^T(x_k)(x - x_k).$$

Preferred choice (larger convergence radius in practise)

$$F_k(x) = f(x) + \sum_{i=1}^m \nu_i^k \bar{f}_i^k(x) + \frac{\mu}{2} \sum_{i=1}^m (\bar{f}_i^k(x))^2$$

- **Unconstraint formulation:** obtain unconstraint formulation using smooth approximation to feasibility set indicator function.

Assignment Project Exam Help

<https://powcoder.com>

Add WeChat powcoder