Numerical Optimisation
Constraint optimisation:
Penalty and augmented Lagrangian methods

f.rullan@cs.uc

Department of Computer Sc
Centre for Medical Image Com
Centre for Inverse Problems
University College London

Lecture 14

Constraint optimization problem

$$\min_{x \in \mathcal{D} \subseteq \mathbb{R}^n} f(x) \qquad \text{(COP)}$$
$$\text{subject to} \quad f_i(x) \quad 0, \quad i = 1, \ldots, m,$$

Con
constraints.

Idea: Minimise a merit function $Q(x;$

Some minimisers of $Q(x;\mu)$ approach tho
constraints as μ approach some set $\mathcal{M}$.

Benefit: reformulation as an unconstraint problem.

Consider a problem with equality constraints

$$\min_{x \in \mathcal{D} \subset \mathbb{R}^n} \quad f(x) \qquad \text{(COP:E)}$$
$$\text{subject to} \quad h_i(x) = 0, \quad i = 1, \dots, p.$$

The m

$$Q(x; \mu) := f(x) + \frac{}{2} \qquad h^2(x),$$

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

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

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

-

- $\nabla h_i(x^\star)$ are linearly independent, then $x^\star$ is a KKT point for (COP:E), and we have that

$$\lim_{k \to \infty} \mu_k h_i(x_k) = \nu_i^\star,$$

where $\nu^\star$ is the multiplier vector that satisfies the KKT conditions for (COP:E).

Proof:

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

From the convergence criterium $_x Q(x_k;\mu_k)$ τ_k (using the ineq

$\overline{\mu_k}$

$|| \; i=1 \quad ||$

As $k \to \infty$: $\tau_k \to 0$, $||\nabla f(x_k)|| \to ||\nabla f$

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

i) If $h_i(x^\star) \neq 0, i = 1, \ldots, p$ then $\nabla h_i(x^\star)$ are linearly dependent which implies that $x^\star$ is a stationary point of $\|h(x)\|^2$.

ii) If $\nabla h_i(x^\star), i = 1, \ldots, p$ are linearly independent, $h_i(x^\star) = 0$ and $x^\star$ 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

Intu

As k

$$\mathcal{L}(x^\star; \nu^\star) = f(x^\star) + \sum_{i}^{p} \tag{2}$$

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

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

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

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

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

$$\ldots_k)].$$

Taki

$$\lim_{k\to\infty} \nu^k = \nu^\star = -(A(x^\star)A(x \ldots$$

and the same in (1) yields the "dual feasibility" con

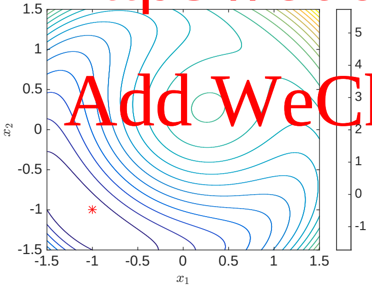
$$\nabla f(x^\star) + A(x^\star)^{\mathrm{T}} \nu^\star = 0.$$

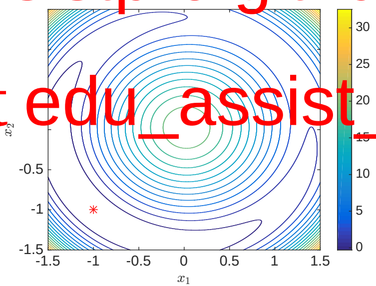
Hence, $x^\star$ is the KKT point with unique Lagrange multiplier $\nu^\star$.

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

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

Qua $\quad - \quad {}_2^2 - 2)^2$.





$$\min \quad -5x_1^2 + x_2^2$$

Solu

Quadratic penalty function: $Q(x;\mu) =$ $\quad ^2 \quad ^2 \quad \frac{\mu}{} \quad ^2.$

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

The iterates would diverge. Unfortunately, a co

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

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

If x i

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

Alternative formulation avoids ill-conditioni

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

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

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

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

Note: Q may be less smooth than the objective a

functions e.g. $c_1(x) = x_1 \geq 0$, then max

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
$(x;\mu_k)$

- There is no guarantee that $\|\nabla_x Q(x;\mu_k)$ τ_k will be satisfied. Practical implementations ne increase μ (and possibly restore the initi constraint violation is not decreasing fast e iterates appear diverging.

- When only equality constraints are present, $Q(x;\mu_k)$ is smooth and algorithms for unconstraint 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

direction due to inadequacy of quadratic approximation.

- Choice of initial point e.g. warm start ... e 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.

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

wher

Let
order necessary conditions with Lagrange mult $\star$ $\star$
$x^\star$ is a local minimiser of $Q_1(x;\mu)$ for all ∞.

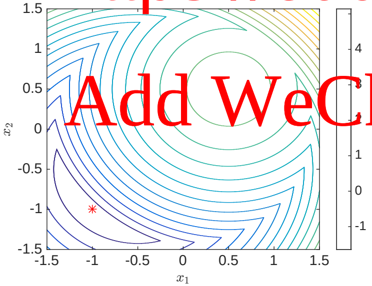
If moreover, the 2nd order sufficient conditions h
then $x^\star$ is a strict local minimiser or Q_1

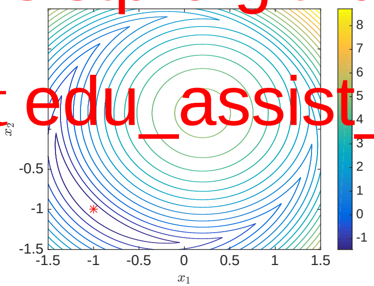
Let $\hat{x}$ be a stationary point of the penalty function $Q_1(x;\mu)$ for all $\mu > \hat{\mu} > 0$. Then, if $\hat{x}$ is feasible for (COP), it satisfies KKT conditions. If $\hat{x}$ is not feasible for (COP), it is an infeasible stationary point.

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

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

ℓ_1 pe





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 fe

Obviously, in the limit $\mu_k \to \infty$, $h_i(x)$
systematic perturbation for moderate values of

Augmented Lagrangian:

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

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

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

Opti

$$0 \approx \nabla_x \mathcal{L}(x_k, \nu^\star) = \nabla f(x_k) \quad \star$$

Comparison yields (an update scheme for

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

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

$$\min \quad x_1 + x_2$$

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

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

Aug

$\mathcal{L}(x,$

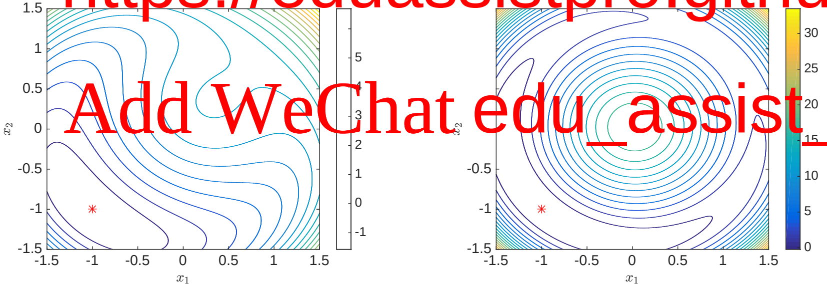


Figure: $\nu = 0.4$

Let $x^\star$ 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 $\nu^\star$. Then for all $\mu \geq \bar{\mu} > 0$, $x^\star$ is a strict local minimiser of $\mathcal{L}_A(x, \nu^\star; \mu)$.

Furthermore, there exist $\delta, \epsilon, M > 0$ such that for all ν^k, μ_k satis

- ϵ, has a unique solution x_k and it holds

$$\|x_k - x^\star\| \leq M\|$$

- it holds

$$\|\nu^{k+1} - \nu^\star\| \leq M\|\nu \;\; - \nu \;\; \|/\mu_k,$$

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

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

gradient algorithm

$$x_{k+1} = P(x_k - \nabla_x \mathcal{L}_A(x$$

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

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

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

  Choose F_k as

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

  Preferred choice (larger convergence rad

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

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