Ch 15: Fundamentals of algorithms for non-linear costrained optimization

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{subject to} \quad \begin{array}{l} c_i(x) = 0, \ i \in \mathcal{E} \\ c_i(x) \geq 0, \ i \in \mathcal{I} \end{array}$$

Generate a sequence of x^* and Lagrange multipliers.

Study 15.1 + 15.2 and return to 15.3–15.6 as needed:

15.1 Categories

(I) Ch 16 Quadratic Problems

- Very efficient with ideas used by sequantial QP, etc.

(II) Ch 17 Combine constr + obj and use unconstr. opt

Eg: For equality constr:

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

with $\mu > 0$, the penalty parameter.

Exact penalty function:

$$f(x) + \mu \sum_{i\in\mathcal{E}} |c_i(x)| \sim (*)$$

for suff. large $\mu > 0$.

However, (*) is non-diff'ble. Attacked using a sequence of smooth subproblems.

Augmented Lagrangian:

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

Here, the basic idea is to fix λ, μ and optimize for x. Then fix x and estimate λ, μ. Repeat.

Will not work well with quadratic $c_i^2(x)$.

Ch 18: SQP

Solve $f(x) + \frac{\mu}{2} \sum_{i \in \mathcal{E}} c_i^2(x)$

using:

$$\min_{\underline{p}} \tfrac{1}{2} p^T \nabla_{xx}^2 \mathcal{L}(x_k, \lambda_k) \underline{p} + \nabla f(x_k)^T \underline{p} \Big\} \text{ Loca quadratic model}$$

subject to:

$$\left.\begin{aligned} \nabla c_i(x_k)^T \underline{p} + c_i(x_k) &= 0, \quad i \in \mathcal{E} \\ \nabla c_i(x_k)^T \underline{p} + c_i(x_k) &\geq 0, \quad i \in \mathcal{I} \end{aligned}\right\} \text{constraint linearization}$$

to move from x_k to $x_k + p$.

Then repeat.

* Can add trust-region and also quasi-Newton approximate Hessians for $\nabla_{xx}^2 \mathcal{L}(x_k, \lambda_k)$.
* Seq Linear-Quadratic programming removes $\nabla_{xx}^2 \mathcal{L}$ and solves an LP problem with trust region first. Second, $\underline{p_k}$ is found by solving equality-constr subproblem.

Ch 19: Interior-point methods for non-linear programming

- Extensions of primal-dual interior point methods for LP (Ch. 14). We can see them as barrier methods by solving:

$$\min_{x,s} \; f(x) - \mu \sum_{i=1}^{m} \log s_i$$

subject to:

$$c_i(x) = 0, \quad i \in \mathcal{E}$$
$$c_i(x) - s_i = 0, \quad i \in \mathcal{I}$$

for $\mu > 0$, a barrier parameter and slack variables $s_i > 0$.

- They compete with SQP and they are new.
- Ch 16, 18, 19 require elimination techniques discussed in 15.3.
- Merit functions and filters help convergence from distant points.

15.2 Combinatorial Difficulty of inequality-constrained problems

Challenge: identifying active constraints.

Guess A^*, the active constraints ... where "$\geq$" is actually "$=$".
Let W denote the guess (working set).

If KKT satisfied with W, then done.
Else pick another W.

$2^{|I|}$ where $|I|$ is the number of inequality constraints.

Example 15.1

$$\min_{x,y} \; f(x,y) \overset{\text{def}}{=} \tfrac{1}{2}(x-2)^2 + \tfrac{1}{2}(y-\tfrac{1}{2})^2$$

Subject to:

$$(x+1)^{-1} - y - \tfrac{1}{4} \geq 0 \quad - 1$$
$$x \geq 0 \quad - 2$$
$$y \geq 0 \quad - 3$$

We have 3 inequality constraints:

$2^3 = 8$ possibilities

$x=0,\ y=0,\ (x+1)^{-1}-y-\frac{1}{4}=0$ — ①

and

$x=0$ — ②

$y=0$ — ③

$x=0,\ y=0$ — ④

$x=0,\ (x+1)^{-1}-y-1/4=0$ — ⑤

$y=0,\ (x+1)^{-1}-y-1/4=0$ — ⑥

No constraints at-all — ⑦

$(x+1)^{-1}-y-\frac{1}{4}=0$ — ⑧

For each case, consider (12.39), and then try to solve the equalities.

15-3 Elimination of variables skipped

15.4 Merit functions and filters

Consider:

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

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

* μ is the penalty parameter
* it is an ℓ_1 function and not diff'ble.
* ϕ_1 is exact ...

Def 15.1 Exact merit function

$\phi(x;\mu)$ is exact if $\exists \mu^*$ s.t. $\mu > \mu^*$ gives that a local soln to the full problem with all constraints will also be a local minimizer of $\phi(x;\mu)$.

For ℓ_1, thm 17.3 gives:

$$\mu^* = \max\{|\lambda_i^*|, i \in \mathcal{E}\cup\mathcal{I}\}$$

where λ_i^* are associated with x^*.

For <u>equality constraints</u>, the Fletcher-augmented Lagrangian is:

$$\boxed{\phi_F(x;\mu) = f(x) - \lambda(x)^T c(x) + \tfrac{1}{2}\mu \sum_{i\in\mathcal{E}} c_i(x)^2}$$

$\boxed{\text{does not suffer from Maratos effect}}$

where $\mu > 0$ and:

$$\lambda(x) = \left[A(x)A(x)^T\right]^{-1} A(x)\nabla f(x)$$

<u>and</u> $A(x) =$ Jacobian of $c(x)$;

$$A(x) = \begin{bmatrix} \nabla c_1(x)^T \\ \nabla c_2(x)^T \\ \vdots \\ \nabla c_m(x)^T \end{bmatrix}$$

The <u>std augmented Lagrangian</u> for equalities:

$$\mathcal{L}_A(x,\lambda;\mu) = f(x) - \lambda^T c(x) + \tfrac{1}{2}\mu \|c(x)\|_2^2$$

... look for sufficient decrease.

Basic ideas:

* Use directional derivatives for ℓ_1, ℓ_2
* Require: for line-search:

$$\phi(x+\alpha p;\mu) \le \phi(x;\mu) + \eta\alpha D(\phi(x;\mu);\underline{p})$$

directional derivative

$\eta \in (0,1)$ and α is step-length.

* Trust-region requires sufficient decrease.

15-10/15

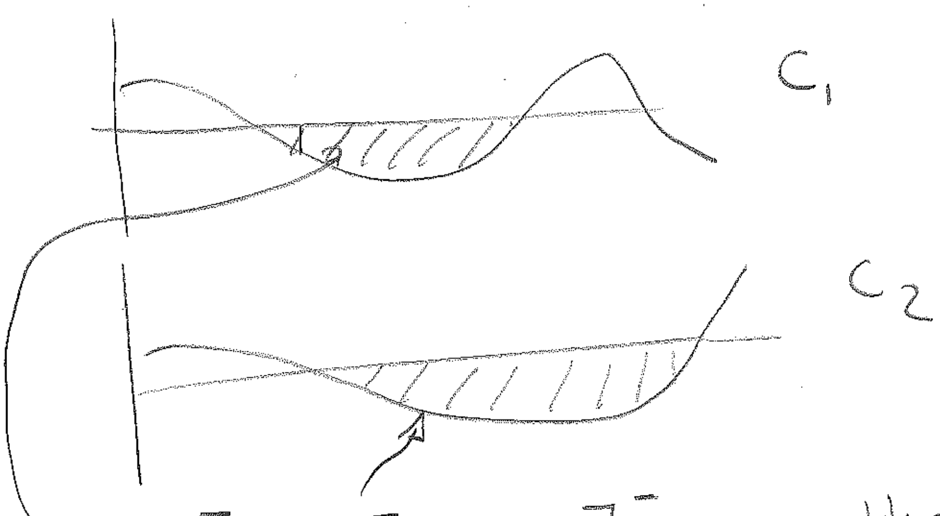
Filters

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

where: $[c_i(x)]^- = \max(-c_i(x), 0) \geq 0$

takes the negative part of the "≥ 0" and tries to minimize it.

Eg:



$[c_1(x)]^- + [c_2(x)]^-$ is the sum of these two functions.

Multi-objective formulation:

Find x s.t. $\min_x f(x)$ and $\min_x h(x)$

Accept x^+ if $(f(x^+), h(x^+))$ is not dominated by other (f_l, h_l).

Def. 15.2

(a) A pair (f_k, h_k) is said to dominate another pair (f_l, h_l) <u>if</u>

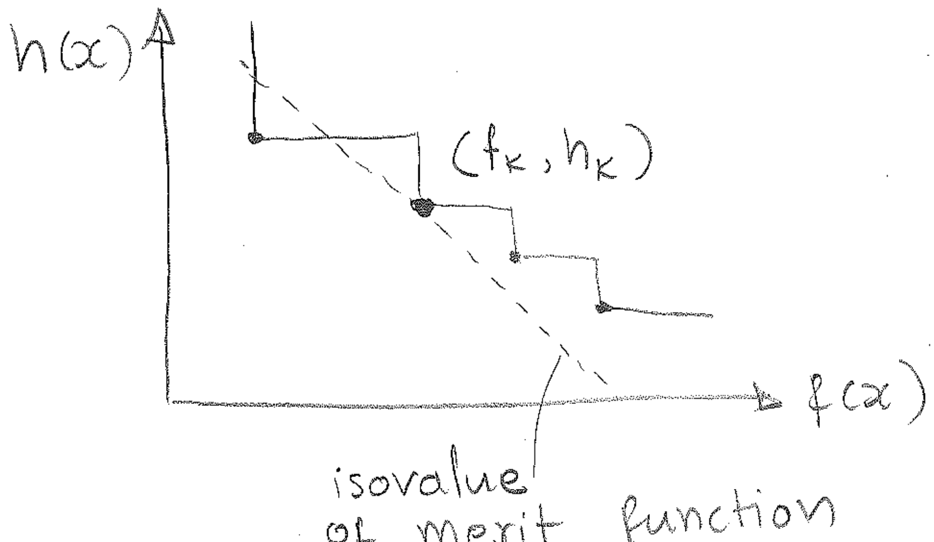
$$f_k \leq f_l \quad \underline{\text{and}}$$
$$h_k \leq h_l$$

(b) A filter is a <u>list</u> of pairs s.t. no pair dominates another one.

(c) An iterate x_k is said to be acceptable to the <u>filter</u> if (f_k, h_k) is not dominated by any pair in the filter.

Basic idea:

* Add (f_k, h_k) to the filter <u>and</u> remove any pairs dominated by it.



We accept x^+ provided:

$$f(x^+) \leq f_J - \beta h_J \quad (*)$$

or $h(x^+) \leq h_J - \beta h_J$

for all existing (f_J, h_J)

where $\beta \in (0,1)$ $(\beta = 10^{-5}$ ok$)$

Better: Replace $(*)$ by:

$$f(x^+) \leq f_J - \beta h^+$$

from (f^+, h^+), the new point.

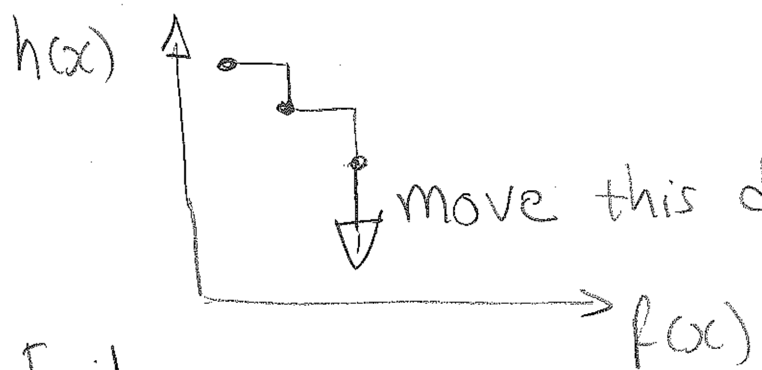
Problem: small step-lengths α_k
in line search
or small trust-regions

Feasibility restoration phase

* Reduce constraint violation by

$$\min_{x} h(x)$$

* Continue until you meet filter constraint:



* Failure in trust-region method addressed this way.

General Filter Method

Choose x_0 and Δ_0.

$k \leftarrow 0$

Repeat until convergence

If step-generation infeasible (small trust-region)

then Compute x_{k+1} using feasibility restoration phase.

else

Compute trial $x^+ = x_k + p_k$

If (f^+, h^+) is acceptable to the filter

then

$x_{k+1} = x^+$ and add (f_{k+1}, h_{k+1}) to the filter

Choose $\Delta_{k+1} \geq \Delta_k$

Remove pairs dominated by (f_{k+1}, h_{k+1})

else Reject x_{k+1} : $x_{k+1} = x_k$

Choose $\Delta_{k+1} < \Delta_k$

end

end

$k \leftarrow k+1$

end repeat

15.5 The Maratos effect

Maratos effect:

* P_k for quadratic convergence may increase both f_k, h_k.

Solutions:

1. If we only have equality constraints, then use:

$$\phi_F(x;\mu) = f(x) - \lambda^T(x)\, c(x) + \tfrac{1}{2}\mu \sum_{i\in\mathcal{E}} c_i(x)^2$$

2. Use second order correction by adding a step $\hat{p}_k$ from $c(x_k + p_k)$

3. Allow merit function to increase for certain iterations.

15.6 Second-order Correction and non-monotone techniquer

Algorithms 15.2 and 15.3 use 2, 3.