Globalization Strategies and Mechanisms

GIAN Short Course on Optimization: Applications, Algorithms, and Computation

Sven Leyffer

Argonne National Laboratory

September 12-24, 2016

Outline

1. Introduction

2. Globalization Strategy: Converge from Any Starting Point
   - Penalty and Merit Function Methods
   - Filter and Funnel Methods
   - Non-Monotone Filter Methods

3. Globalization Mechanisms
   - Line-Search Methods
   - Trust-Region Methods

Recap: Methods for Nonlinear Optimization

Considered three classes of methods

Sequential Quadratic Programming (SQP)

- Solve sequence of QP approximations
- Similar to Newton's method ... may fail

Interior-Point Methods (IPM)

- Solve sequence of perturbed KKT systems
- Perturbation of Newton's method ... may fail

Augmented Lagrangian Methods

- Approx. minimize augmented Lagrangian
- Converge ... but assumptions really strong

Add Global Convergence Mechanisms

Mechanism should interfere as little as possible with method.

Motivation

(NLP) $\underset{x}{\text{minimize}} \; f(x) \quad \text{subject to } c(x) = 0 \quad x \geq 0$

Local methods (e.g. SQP) may not converge if started far from x^*
... barrier methods require unrealistic assumptions (global solve)

Equip local methods with globalization strategy and mechanism
... to ensure convergence from remote starting points

Globalization Strategy

- How do we decide a point is better?
- Uniike unconstrained case, balance objective and feasibility

Globalization Mechanism

- Generalize line-search or trust-region from unconstrained case

Outline

General Outline of Globalization Strategy

$$\text{(NLP)} \qquad \begin{cases} \underset{x}{\text{minimize}} & f(x) \\ \text{subject to} & c(x) = 0 \\ & x \geq 0, \end{cases}$$

Goal and Limitations

Ensure convergence from remote starting points,
i.e. global convergence $\neq$ global minimum

- Monitor progress of iterates, $x^{(k)}$
- Cannot just use objective decreaseas $f(x^{(k)} + \alpha s^{(k)}) < f(x^{(k)})$
- Must also look at constraint violation, e.g. $\|c(x)\|$

Penalty and Merit Function Methods

$$\text{(NLP)} \qquad \underset{x}{\text{minimize}} \; f(x) \quad \text{subject to } c(x) = 0, \; x \geq 0$$

Combine objective and constraints, e.g. exact penalty function

$$p_\rho(x) = f(x) + \rho \|c(x)\|,$$

where $\rho > 0$ is penalty parameter

- Local minimizers of $p_\rho(x)$ are local mins. of (NLP)
- Apply unconstrained globalization techniques
- Popular penalty functions: ℓ_1 and ℓ_2 penalty functions

Theorem (Equivalence of Local Minimizers)

If the penalty parameter is sufficiently large, i.e. $\rho > \|y^*\|_D$, then a local minimizers of $p_\rho(x)$ is a local min of (NLP).

- y^* optimal multiplier corresponding to x^*
- $\|\cdot\|_D$ is the dual e.g. ℓ_∞-norm is dual of ℓ_1-norm
- Monitor progress of SQP, IPM methods using penalty function

Penalty and Merit Function Methods

$$(\text{NLP}) \qquad \underset{x}{\text{minimize}} \; f(x) \quad \text{subject to } c(x) = 0, \; x \geq 0$$

Nonsmooth penalty function (e.g. ℓ_1-norm)

$$\underset{x}{\text{minimize}} \; p_\rho(x) = f(x) + \rho \|c(x)\|_1$$

Can formulate equivalent smooth problem

$$(\text{NLP}) \qquad \begin{cases} \underset{x}{\text{minimize}} & f(x) + \rho \sum_{i=1}^{m} \left(s_i^+ + s_i^-\right) \\ \text{subject to} & c(x) = s^+ - s^- \\ & x \geq 0, \; s^+ \geq 0, \; s^- \geq 0 \end{cases}$$

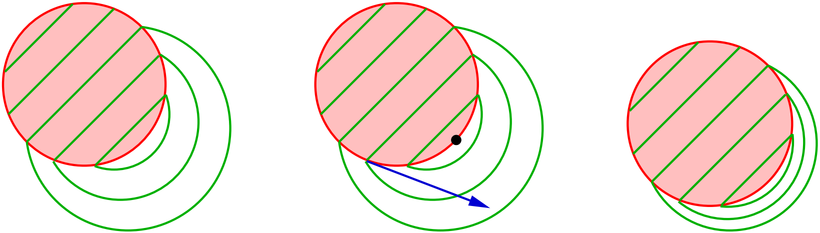
... apply SQP to this problem

ℓ_1 Exact Penalty Function & Maratos Effect

$$\underset{x}{\text{minimize}} \; p(x;\rho) \;=\; f(x) + \rho\|c(x)\|_1 \quad \text{subject to } x \geq 0$$

where $\|c(x)\|_1$ constraint violation

- $p(x;\rho)$ nonsmooth, but equivalent to smooth problem
- Penalty parameter not known a priori: $\rho > \|y^*\|_\infty$
- Large penalty parameter $\Rightarrow$ slow convergence; inefficient



Maratos effect motivates second-order correction steps

Filter Methods for Global Convergence

Provide alternative to penalty methods

- Optimal penalty parameter, $\rho > \|y^*\|_D$ not known a priori
  - Penalty adjustment can be problematc ... avoid $\rho_k \to \infty$
  - Modern methods solve two subproblems (LP and QP) to adjust ρ_k
- Poor practical convergence, if ρ_k large for highly nonlinear constraints

View penalty function as two competing aims:

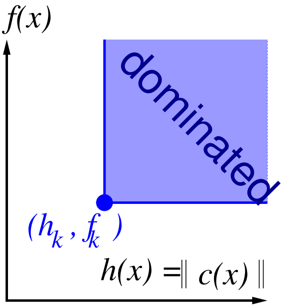
1. Minimize $f(x)$
2. Minimize $h(x) := \|c(x)\|$ ... more important

... borrow ideas from multi-objectiv eoptimization

Filter Methods for NLP

Penalty function combines two competing aims:

1. Minimize $f(x)$
2. Minimize $h(x) := \|c^-(x)\|$ ... more important



Borrow concept of domination from multi-objective optimization

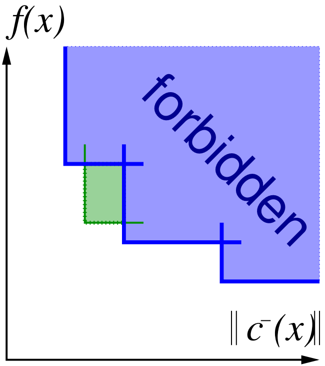
$$(h^{(k)}, f^{(k)}) \text{ dominates } (h^{(l)}, f^{(l)})$$
$$\text{iff} \quad h^{(k)} \leq h^{(l)} \;\&\; f^{(k)} \leq f^{(l)}$$

i.e. $x^{(k)}$ at least as good as $x^{(l)}$

Filter Methods for NLP

Filter $\mathcal{F}$: list of non-dominated pairs $(h^{(l)}, f^{(l)})$

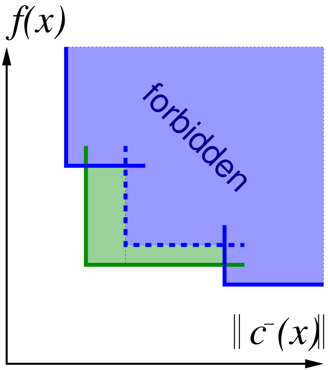
- new $x^{(k+1)}$ acceptable to filter $\mathcal{F}$, iff
  1. $h^{(k+1)} \leq h^{(l)} \; \forall l \in \mathcal{F}$, or
  2. $f^{(k+1)} \leq f^{(l)} \; \forall l \in \mathcal{F}$



Filter Methods for NLP

Filter $\mathcal{F}$: list of non-dominated pairs $(h^{(l)}, f^{(l)})$

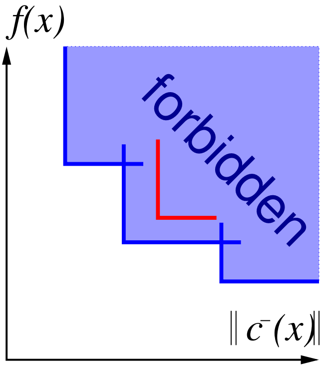
- new $x^{(k+1)}$ acceptable to filter $\mathcal{F}$, iff
  1. $h^{(k+1)} \leq h^{(l)} \; \forall l \in \mathcal{F}$, or
  2. $f^{(k+1)} \leq f^{(l)} \; \forall l \in \mathcal{F}$
- remove redundant entries



Filter Methods for NLP

Filter $\mathcal{F}$: list of non-dominated pairs $(h^{(l)}, f^{(l)})$

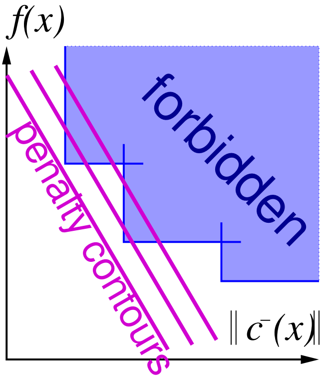
- new $x^{(k+1)}$ acceptable to filter $\mathcal{F}$, iff
  1. $h^{(k+1)} \leq h^{(l)} \ \forall l \in \mathcal{F}$, or
  2. $f^{(k+1)} \leq f^{(l)} \ \forall l \in \mathcal{F}$
- remove redundant entries
- reject new $x^{(k+1)}$, if $h^{(k+1)} > h^{(l)}$ & $f^{(k+1)} > f^{(l)}$ & reduce trust region $\Delta = \Delta/2$



Filter Methods for NLP

Filter $\mathcal{F}$: list of non-dominated pairs $(h^{(l)}, f^{(l)})$

- new $x^{(k+1)}$ acceptable to filter $\mathcal{F}$, iff
  1. $h^{(k+1)} \leq h^{(l)} \; \forall l \in \mathcal{F}$, or
  2. $f^{(k+1)} \leq f^{(l)} \; \forall l \in \mathcal{F}$
- remove redundant entries
- reject new $x^{(k+1)}$, if $h^{(k+1)} > h^{(l)}$ & $f^{(k+1)} > f^{(l)}$ & reduce trust region $\Delta = \Delta/2$



$\Rightarrow$ often accept new $x^{(k+1)}$, even if penalty function increases

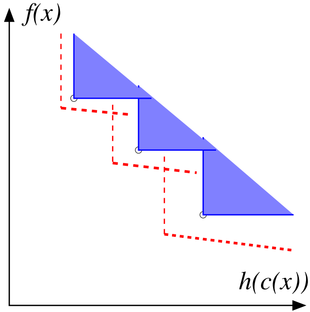
Formal Definition of Step Acceptance

New $x^{(k+1)}$ acceptable
iff either of

1. $h^{(k+1)} \le \beta h^{(l)}$, or
2. $f^{(k+1)} + \gamma h^{(k+1)} \le f^{(l)}$

hold $\forall l \in \mathcal{F}_k$

Lemma: ∞-sequence in $\mathcal{F} \Rightarrow h^{(k)} \to 0$



Sufficient objective reduction:

if predicted reduction $\Delta q^{(k)} > 0$ **then**

check $f(x^{(k)}) - f(x^{(k+1)}) \ge \sigma \, \Delta q^{(k)}$

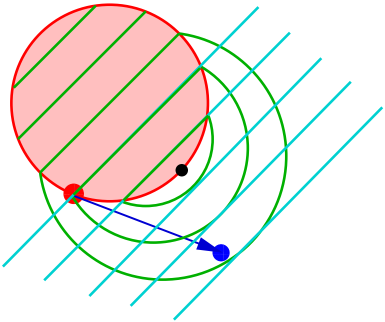
where $\Delta q^{(k)} = {g^{(k)}}^T s + \frac{1}{2} s^T H^{(k)} s$

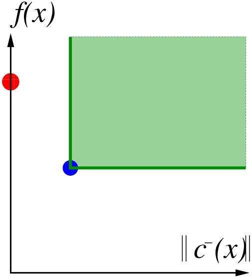
Constants: $\beta = 0.999, \gamma = 0.001, \sigma = 0.1$

The Maratos Example Revisited

Filter methods work well for Maratos example ...

1. Maratos step decreases objective & increases constraints
2. Maratos step acceptable to filter





More Filter Methods

1. IPOPT free interior-point line-search filter method
   - [Wächter & Biegler, 2005] (3 papers on theory & results)
   - tighter "switching condition" & 2nd-order correction steps
     $\Rightarrow$ superlinear convergence
   - proof is very complicated, not intuitive

2. [S. Ulbrich, 2003] shows second-order convergence
   - surprisingly: no 2nd-order correction steps
   - replace $f(x)$ in filter by Lagrangian:
     $\mathcal{L}(x, y, z) := f(x) - y^T c(x) - z^T x$
   - replace $\|c(x)\|$ in filter by $\|c(x)\| + z^T x$
   - modify "switching condition" & feasibility

More Filter Methods

3. Pattern search filter [Audet & Dennis, 2000]
   - filter plus one feasible iterate x_F: $f(x^{(k+1)}) < f(x_F)$
   - only require decrease; no sufficient reduction
   - converges to x_* where "$0 \in \partial f(x_*)$" or "$0 \in \partial \|c(x_*)\|$"
     $\Rightarrow$ convergence to "KKT points"???
4. Nonsmooth bundle-filter:
   - [Lemaréchal et al, 1995] convex hull of filter points
   - [Fletcher & L, 1999] straightforward extension of NLP
   - [Karas et al, 2006] "improvement function" & filter ???
5. Filter for nonlinear complementarity [Nie, 2005]
6. Filter for Genetic Algorithms ... standard technique
7. Filter methods for feasibility restoration: min $\|c^-(x)\|$

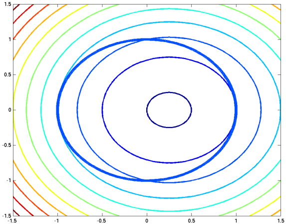
Removing the Need for Second-Order Corrections

Filter methods also suffer from Maratos Effect:

$$\begin{array}{l} \text{minimize} \;\; 2(x_1^2 + x_2^2 - 1) - x_1 \\ \text{subject to } x_1^2 + x_2^2 - 1 = 0 \end{array}$$

... example due to Conn, Gould & Toint

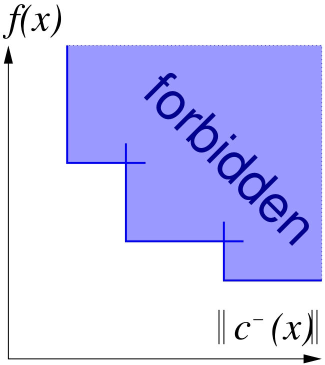
Start x_0 near $(1, 0)$
$\Rightarrow f_1 > f_0$ and $h_1 > h_0$ reject
$\Rightarrow$ need second-order correction (SOC) steps
SOC steps are cumbersome
... can we avoid them?
Idea: Use non-monotone filter ...



Idea of Non-Monotone Filter

Consider Shadow Filter:

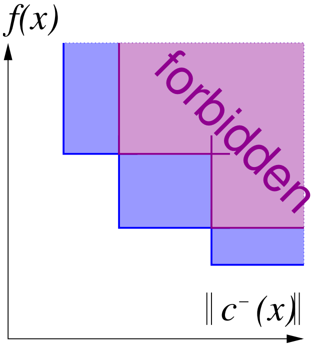
- accept new point $x^{(k+1)}$, if dominated by less than $M \geq 0$ filter entries
- standard filter: $M = 0$
- filter $\simeq$ semi-permeable membrane
- count dominating entries



Idea of Non-Monotone Filter

Consider Shadow Filter:

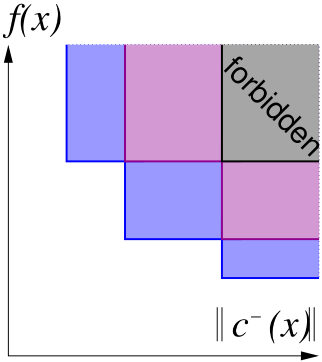
- accept new point $x^{(k+1)}$,
  if dominated by less than $M \geq 0$
  filter entries
- standard filter: $M = 0$
- filter $\simeq$ semi-permeable membrane
- count dominating entries



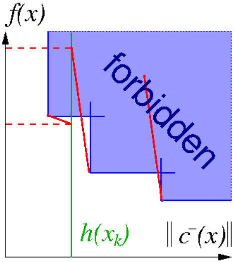
Idea of Non-Monotone Filter

Consider Shadow Filter:

- accept new point $x^{(k+1)}$,
  if dominated by less than $M \geq 0$
  filter entries
- standard filter: $M = 0$
- filter $\simeq$ semi-permeable membrane
- count dominating entries



Non-Monotone Sufficient Reduction Test



- Similar unconstrained optimization
- Actual reductn $\geq$ predicted reductn: $f(x^{(k)}) - f(x^{(k)} + s) \geq \sigma \Delta q^{(k)}$ replaced by

$$\left(\max_{i \in \{0,\ldots,M\}} \tilde{f}^{(k-i)} \right) - f(x^{(k)} + s) \geq \sigma \Delta q^{(k)}$$

where for all $(k - i) \in \mathcal{F}_k$

$$\tilde{f}^{(k-i)} = \begin{cases} f^{(k-i)} + (h^{(k-i)} - h) * 1000 & \text{if } h^{(k-i} \\ f^{(k-i)} + (h^{(k-i)} - h)/1000 & \text{if } h^{(k-i} \end{cases}$$

- Sufficient decrease after at most M steps
- $M = 0$: monotone reduction

FASTr: A New Nonmonotone Filter Method

Comparing solvers on 410 small CUTEr problems

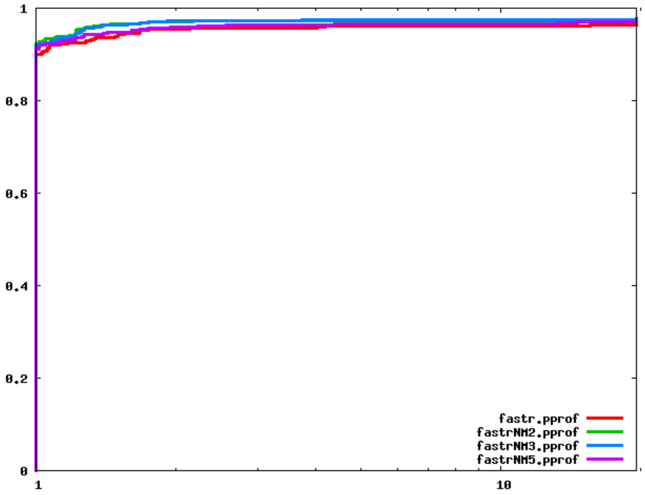
- `FilterSQP` written in `fortran` dates back to 1998
- `FASTr` currently being developed in `C`
  2500 lines of C-code (vs. 5300 lines of `fortran`):
  - Restoration phase re-uses main loop!
  - No second-order correction steps

Performance profiles [Dolan and Moré, 2002]:

$$\forall \;\; \text{solver} \;\; s \quad \text{perf}_s(p) \; := \; \log_2 \left(\frac{\# \; \text{iter}(s,p)}{\text{best\_iter}(p)} \right) \;, \; p \in \text{problem}$$

- Sort in ascending order (step-function)
- Probability that solver s at most 2^x times worse

FASTr(0), FASTr(2), FASTr(3), FASTr(5)



Outline

Globalization Mechanisms

Key algorithmic ingredients

1. Efficient step computation
   e.g. SQP, SLP, IPM, ...
2. Global convergence strategy
   e.g. penalty or filter ...
3. Global convergence mechanism
   ... enforce global strategy

Two Main Global Convergence Mechanisms

1. Line-search methods
2. Trust-region methods

... already reviewed in unconstrained lectures

Line-Search Methods

Given direction, $s^{(k)}$, backtrack along $s^{(k)}$ to acceptable point

Search Directions

1. Interior-point methods use primal-dual direction $s = (\Delta_x, \Delta_y, \Delta_z)$
2. SQP methods obtain search direction from solution of QP

Search direction must be descend direction for penalty function

$$\nabla p(x^{(k)}; \rho)^T s < 0$$

... step computation can ensure descend, e.g. modifying Hessian

Armijo Line-Search Method for NLP

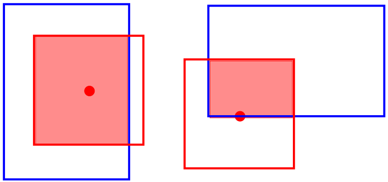
Given $x^{(0)} \in \mathbb{R}^n$, let $0 < \sigma < 1$, set $k = 0$
while $x^{(k)}$ *is not optimal* **do**
  Approx. step computation subproblem around $x^{(k)}$ for s.
  Ensure descend, e.g. $\nabla p(x^{(k)}; \rho)^T s < 0$
  Set $\alpha^0 = 1$ and $l = 0$
  repeat
    Set $\alpha^{l+1} = \alpha^l / 2$ and evaluate $p(x^{(k)} + \alpha^{l+1} s; \rho)$.
    Set $l = l + 1$.
  until $p(x^{(k)} + \alpha^l s; \rho) \leq f^{(k)} + \alpha^l \sigma s^T \nabla p^{(k)}$;
  Set $k = k + 1$.
end

... similar for filter methods

Trust-Region Methods

Trust-region methods restrict step during subproblem

- Add step restriction $\|d\| \leq \Delta_k$ to approximate subproblem
  - Preferred ℓ_2 norm in unconstrained case
  - Prefer ℓ_∞ norm in constrained case
    ... easy to intersect TR with bounds



  - Adjust TR radius Δ_k as before
- Require more effort to compute step
- Have slightly stronger convergence properties

Trust-Region Algorithm Framework

Given $x^{(0)} \in \mathbb{R}^n$, choose $\Delta_0 \geq \underline{\Delta} > 0$, set $k = 0$

repeat

  Reset $\Delta_{k,l} := \Delta^{(k)} \geq \underline{\Delta} > 0$; set `success = false`, and $l = 0$

  repeat

    Solve approx. subproblem in $\|d\| \leq \Delta_{k,l}$

    if $x^{(k)} + d$ *is sufficiently better than* $x^{(k)}$ **then**

      Accept step: $x^{(k+1)} = x^{(k)} + d$; increase $\Delta_{k,l+1}$

      Set `success = true`.

    else

      Reject step decrease TR radius, e.g. $\Delta_{k,l+1} = \Delta_{k,l}/2$.

    end

  until `success = true`;

  Set $k = k + 1$.

until $x^{(k)}$ *is optimal*;

Teaching Points and Summary

Key algorithmic ingredients

1. Efficient step computation
   e.g. SQP, SLP, IPM, ...
2. Global convergence strategy
   e.g. penalty or filter ...
3. Global convergence mechanism
   ... line-search or TR

- Maratos effect prevents Newton steps from being accepted.
- Nonmonotone methods avoid Maratos effect