Active-Set Methods for Large-Scale NLP

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

Sven Leyffer

Argonne National Laboratory

September 12-24, 2016

Outline

Active Set Methods for Nonlinear Programming (NLP)

Nonlinear Program (NLP)

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

where f, c twice continuously differentiable

Definition (Active Set)

Active set: $\mathcal{A}(x) = \{i \mid x_i = 0\}$

Inactive set: $\mathcal{I}(x) = \{1, \ldots, n\} - \mathcal{A}(x)$

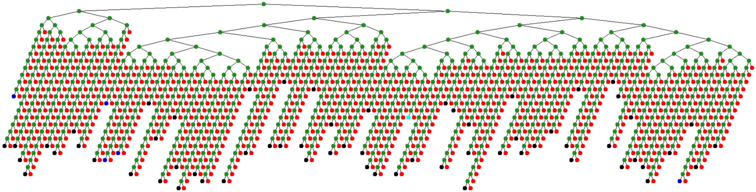
For known optimal active set $\mathcal{A}(x^*)$, just use Newton's method

Goal: develop robust, fast, parallelizable active-set methods

Active Set Methods for Nonlinear Programming (NLP)

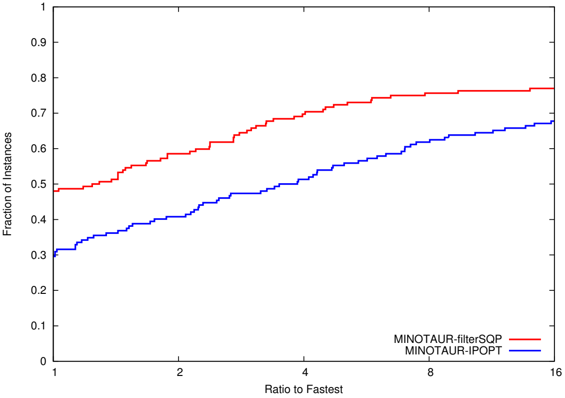
Motivation: mixed-integer nonlinear optimization: $x_i \in \{0, 1\}$

- solve NLP relaxation $x_i \in [0, 1]$
- branch on $\hat{x}_i \notin \{0, 1\}$ ... two new NLPs: $x_i = 0$ or $x_i = 1$
- solve sequence of closely related NLPs



Branch-and-bound solves millions of related NLPs ...

Active-Set vs. Interior-Point Solvers in MINLP



MINOTAUR with FilterSQP vs IPOPT: CPU time

- FilterSQP warm-starts much faster than IPOPT
- similar results for BONMIN (IBM/CMU) solver

It's only half bull

Large-Scale Active-Set Methods

Goal of This Lecture

Towards scalable active-set methods for nonlinear optimization

Two Approaches for Large-Scale Active-Set Methods

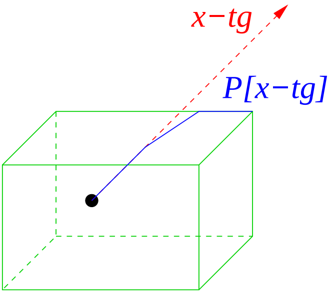
1. Develop methods for large-scale QP
   - Use within SQP framework $\Rightarrow$ readily understood
   - Should work for indefinite QPs
   - Must allow large changes to active set ... not pivoting
   - Should be scalable ... i.e. matrix-free linear algebra
2. Develop methods for large-scale NLP
   - Ensure suitable for matrix-free linear algebra
   - Properly combine step-computation and globalization strategy

Outline

Recall: Projected Gradient for Box Constrained QPs

Simpler box constrained QP ...

$$\begin{aligned} &\underset{x}{\text{minimize}} \quad \tfrac{1}{2}x^T Hx + g^T x \;=:\; q(x) \\ &\text{subject to } l \le x \le u \end{aligned}$$



Projected steepest descent $P[x - \alpha \nabla q(x)]$

- piecewise linear path
- large changes to $\mathcal{A}$-set
  ... but slow (steepest descent)

x^c Cauchy point $\equiv$ first minimum of $q(x(\alpha))$, for $\alpha \ge 0$

Theorem: Cauchy points converge to stationary point.

Projected Gradient & CG for Box Constrained QPs

x^0 given such that $l \leq x^0 \leq u$; set $k = 0$

WHILE (not optimal) **BEGIN**

1. find Cauchy point x_k^c & active set $\mathcal{A}(x_k^c) := \{i | [x_k^c]_i = l_i \text{ or } u_i\}$
2. (approx.) solve box QP in subspace $\mathcal{I} := \{1, \ldots, n\} - \mathcal{A}(x_k^c)$

$$\begin{array}{ll} \underset{x}{\text{minimize}} & \frac{1}{2} x^T H x + g^T x \\ \text{subject to } l \leq x \leq u & \\ \quad x_i = [x_k^c]_i \ , \forall\, i \in \mathcal{A}(x_k^c) & \end{array} \quad \Leftrightarrow \quad \begin{array}{l} \text{apply CG to ...} \\ H_{\mathcal{I},\mathcal{I}} x_{\mathcal{I}} = \ldots \end{array}$$

for x^{k+1}; set $k = k + 1$

END

Cauchy point $\Rightarrow$ global convergence ... but faster due to CG

How to Include $A^T x = b$?

Projection onto box is easy, but tough for general QP

$$P_{QP}[z] = \begin{cases} \underset{x}{\text{minimize}} & (x-z)^T(x-z) \\ \text{subject to} & A^T x = b \\ & l \le x \le u \end{cases}$$

... as hard as original QP! ... Idea: project onto box only

$\Rightarrow$ subspace solve $H_{\mathcal{I},\mathcal{I}} x_{\mathcal{I}} = \ldots$ becomes solve with KKT system

$$\begin{bmatrix} H_{\mathcal{I},\mathcal{I}} & -A_{:,\mathcal{I}} \\ A_{:,\mathcal{I}}^T & \end{bmatrix} \begin{pmatrix} x_{\mathcal{I}} \\ y \end{pmatrix} = \ldots$$

Which gradient / merit function in Cauchy step?

The Augmented Lagrangian

[Arrow & Solow:58], [Hestenes:69], [Powell:69]

$$\boxed{\underset{x}{\text{minimize}} \;\; L(x, y_k, \rho_k) \;=\; f(x) \;-\; y_k^T c(x) + \tfrac{1}{2}\rho_k \|c(x)\|^2}$$

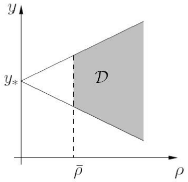
As $y_k \to y_*$:
- $x_k \to x_*$ for $\rho_k > \bar{\rho}$
- No ill-conditioning, improves convergence rate

- An old idea for nonlinear constraints ... smooth merit function
- Poor experience with LPs (e.g., MINOS vs. LANCELOT)
- But special structure of LPs (and QPs) not fully exploited

$$\boxed{f(x) = \tfrac{1}{2}x^T Hx + g^T x \quad \& \quad c(x) = A^T x - b}$$

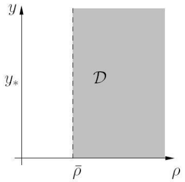
Augmented Lagrangian for Linear Constraints

Nonlinear $c(x)$



$(y_k, \rho_k) \in \mathcal{D}$

$c(x) = A^T x - b$



$\rho_k > \bar{\rho}$

$\forall(\rho, y) \in \mathcal{D}$, minimize $L(x, y, \rho)$ has unique solution $x(y, \rho)$:

- bound constrained augmented Lagrangian converges
- Hessian $\nabla^2_{xx} L(x, y, \rho)$ is positive definite on optimal face
- $\bar{\rho} \approx 2\|H_*\| \;/\; \|A_* A_*\| \Rightarrow$ convergence

QP by Projected Augmented Lagrangian QPPAL

WHILE (not optimal) **BEGIN**

1. Find $\omega_k \searrow 0$ optimal solution x_k^c of

$$\underset{l \le x \le u}{\text{minimize}} \ \tfrac{1}{2}x^T Hx + g^T x - y^T(A^T x - b) + \tfrac{1}{2}\rho_k \|A^T x - b\|^2$$

2. Find $\mathcal{A}(x_k^c)$ & estimate penalty $\bar{\rho} = 2\,\|H_{\mathcal{I}}\|/\|A_{\mathcal{I}} A_{\mathcal{I}}^T\|$
3. **IF** $\bar{\rho} > \rho_k$ **THEN** update $\rho_{k+1} = \bar{\rho}$ & **CYCLE**
   ELSE update multiplier: $y_k^c = y_k - \rho_k(A^T x_k^c - b)$

QP by Projected Augmented Lagrangian QPPAL

WHILE (not optimal) **BEGIN**

1. Find $\omega_k \searrow 0$ optimal solution x_k^c of

$$\underset{l \le x \le u}{\text{minimize}} \ \frac{1}{2}x^T Hx + g^T x - y^T(A^T x - b) + \frac{1}{2}\rho_k \|A^T x - b\|^2$$

2. Find $\mathcal{A}(x_k^c)$ & estimate penalty $\bar{\rho} = 2\,\|H_{\mathcal{I}}\|/\|A_{\mathcal{I}}A_{\mathcal{I}}^T\|$

3. **IF** $\bar{\rho} > \rho_k$ **THEN** update $\rho_{k+1} = \bar{\rho}$ & **CYCLE**
   ELSE update multiplier: $y_k^c = y_k - \rho_k(A^T x_k^c - b)$

4. Solve equality QP in subspace $\rightarrow (\Delta x_{\mathcal{I}}, \Delta y)$

$$\begin{bmatrix} H_{\mathcal{I},\mathcal{I}} & -A_{:,\mathcal{I}} \\ A_{:,\mathcal{I}}^T & \end{bmatrix} \begin{pmatrix} \Delta x_{\mathcal{I}} \\ \Delta y \end{pmatrix} = -\begin{pmatrix} [\nabla x L(x_k^c, y_k^c, 0)]_{\mathcal{I}} \\ A^T x_k^c - b \end{pmatrix}$$

QP by Projected Augmented Lagrangian QPPAL

WHILE (not optimal) **BEGIN**

1. Find $\omega_k \searrow 0$ optimal solution x_k^c of

$$\underset{l \le x \le u}{\text{minimize}} \; \frac{1}{2} x^T H x + g^T x - y^T (A^T x - b) + \frac{1}{2} \rho_k \|A^T x - b\|^2$$

2. Find $\mathcal{A}(x_k^c)$ & estimate penalty $\bar{\rho} = 2 \, \|H_{\mathcal{I}}\| / \|A_{\mathcal{I}} A_{\mathcal{I}}^T\|$
3. **IF** $\bar{\rho} > \rho_k$ **THEN** update $\rho_{k+1} = \bar{\rho}$ & **CYCLE**
   ELSE update multiplier: $y_k^c = y_k - \rho_k (A^T x_k^c - b)$
4. Solve equality QP in subspace $\rightarrow \; (\Delta x_{\mathcal{I}}, \Delta y)$

$$\begin{bmatrix} H_{\mathcal{I},\mathcal{I}} & -A_{:,\mathcal{I}} \\ A_{:,\mathcal{I}}^T & \end{bmatrix} \begin{pmatrix} \Delta x_{\mathcal{I}} \\ \Delta y \end{pmatrix} = - \begin{pmatrix} [\nabla x L(x_k^c, y_k^c, 0)]_{\mathcal{I}} \\ A^T x_k^c - b \end{pmatrix}$$

5. Line-search on $L(x_k^c + \alpha \Delta x, y_k^c + \alpha \Delta y, \rho)$; **update** x, y, k, ρ

END

QP by Projected Augmented Lagrangian QPPAL

WHILE (not optimal) **BEGIN**

1. Find $\omega_k \searrow 0$ optimal solution x_k^c of

$$\underset{l \le x \le u}{\text{minimize}} \ \frac{1}{2}x^T Hx + g^T x - y^T(A^T x - b) + \frac{1}{2}\rho_k \|A^T x - b\|^2$$

2. Find $\mathcal{A}(x_k^c)$ & estimate penalty $\bar{\rho} = 2\,\|H_{\mathcal{I}}\|/\|A_{\mathcal{I}} A_{\mathcal{I}}^T\|$
3. **IF** $\bar{\rho} > \rho_k$ **THEN** update $\rho_{k+1} = \bar{\rho}$ & **CYCLE**
   ELSE update multiplier: $y_k^c = y_k - \rho_k(A^T x_k^c - b)$
4. Solve equality QP in subspace $\rightarrow$ $(\Delta x_{\mathcal{I}}, \Delta y)$

$$\begin{bmatrix} H_{\mathcal{I},\mathcal{I}} & -A_{:,\mathcal{I}} \\ A_{:,\mathcal{I}}^T & \end{bmatrix} \begin{pmatrix} \Delta x_{\mathcal{I}} \\ \Delta y \end{pmatrix} = -\begin{pmatrix} [\nabla x L(x_k^c, y_k^c, 0)]_{\mathcal{I}} \\ A^T x_k^c - b \end{pmatrix}$$

5. Line-search on $L(x_k^c + \alpha\Delta x, y_k^c + \alpha\Delta y, \rho)$; **update** x, y, k, ρ

END

1.-3. identify active set while 4. gives fast convergence

Forcing Sequences of Augmented Lagrangian Methods

In general, two competing aims in augmented Lagrangian:

1. reduce $h_k := \|A^T x_k - b\| \le \eta_k \searrow 0$
2. reduce $\theta_k := \|\nabla L(x_k, y_k, \rho_k) - z_k\| \le \omega_k \searrow 0$

Note: QPPAL does not need η_k,
see also [Dostal:99], [Delbos&Gilbert:03]

Forcing Sequences of Augmented Lagrangian Methods

In general, two competing aims in augmented Lagrangian:

1. reduce $h_k := \|A^T x_k - b\| \leq \eta_k \searrow 0$
2. reduce $\theta_k := \|\nabla L(x_k, y_k, \rho_k) - z_k\| \leq \omega_k \searrow 0$

Note: QPPAL does not need η_k,
see also [Dostal:99], [Delbos&Gilbert:03]

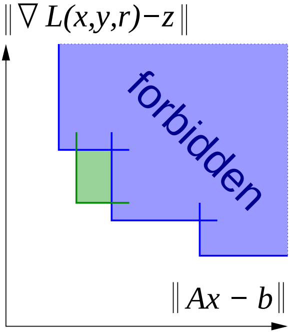
... two arbitrary sequences, η_k, ω_k

... why should one sequence $\{\omega_k\}, \{\eta_k\}$ fit *all problems ???*

A Filter for Augmented Lagrangian Methods

Introduce a filter $\mathcal{F}$ for convergence

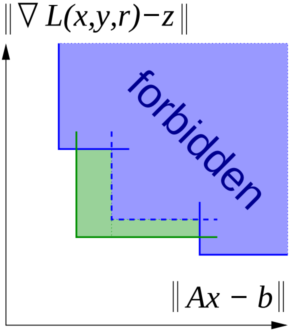
- list of pairs $(\|A^T x_l - b\|, \|\nabla L_l - z_l\|)$
- no pair dominates any other pair
- new x_k acceptable to filter $\mathcal{F}$, iff
  1. $h_k \leq 0.99 \cdot h_l \; \forall l \in \mathcal{F}$
  2. $\theta_k \leq 0.99 \cdot \theta_l \; \forall l \in \mathcal{F}$



A Filter for Augmented Lagrangian Methods

Introduce a filter $\mathcal{F}$ for convergence

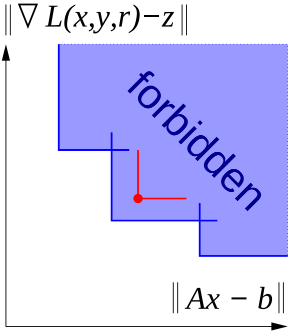
- list of pairs $(\|A^T x_l - b\|, \|\nabla L_l - z_l\|)$
- no pair dominates any other pair
- new x_k acceptable to filter $\mathcal{F}$, iff
  1. $h_k \leq 0.99 \cdot h_l \; \forall l \in \mathcal{F}$
  2. $\theta_k \leq 0.99 \cdot \theta_l \; \forall l \in \mathcal{F}$
- remove redundant entries



A Filter for Augmented Lagrangian Methods

Introduce a filter $\mathcal{F}$ for convergence

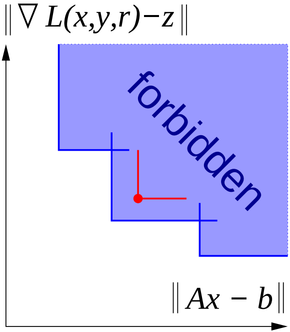
- list of pairs $(\|A^T x_l - b\|, \|\nabla L_l - z_l\|)$
- no pair dominates any other pair
- new x_k acceptable to filter $\mathcal{F}$, iff
  1. $h_k \leq 0.99 \cdot h_l \; \forall l \in \mathcal{F}$
  2. $\theta_k \leq 0.99 \cdot \theta_l \; \forall l \in \mathcal{F}$
- remove redundant entries
- reject new x_k, if $h_k \geq h_l$ & $\theta_k \geq \theta_l$



A Filter for Augmented Lagrangian Methods

Introduce a filter $\mathcal{F}$ for convergence

- list of pairs $(\|A^T x_l - b\|, \|\nabla L_l - z_l\|)$
- no pair dominates any other pair
- new x_k acceptable to filter $\mathcal{F}$, iff
  1. $h_k \leq 0.99 \cdot h_l \; \forall l \in \mathcal{F}$
  2. $\theta_k \leq 0.99 \cdot \theta_l \; \forall l \in \mathcal{F}$
- remove redundant entries
- reject new x_k, if $h_k \geq h_l$ & $\theta_k \geq \theta_l$



... an old friend from Chicago (Castellmarre di Stabia)

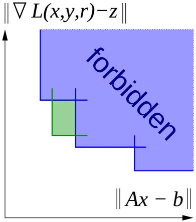
Augmented Lagrangian Cauchy Pointe (Al Capone)

Requirement on Cauchy Point x_k^c for filter:

1. x_k^c, y_k^c acceptable to filter
2. $\|\nabla L(x_k, y_k, \rho_k) - z_k\| \leq \omega_k$
   ... optimality of Lagrangian

Augmented Lagrangian Cauchy Pointe (Al Capone)

Requirement on Cauchy Point x_k^c for filter:



1. x_k^c, y_k^c acceptable to filter
2. $\|\nabla L(x_k, y_k, \rho_k) - z_k\| \leq \omega_k$
   ... optimality of Lagrangian

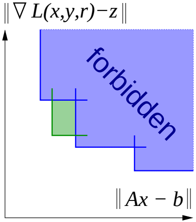
   New: $\omega_k := 0.1 \max \{\|\nabla L_l - z_l\|\}$
   ... depends on filter

Augmented Lagrangian Cauchy Pointe (Al Capone)

Requirement on Cauchy Point x_k^c for filter:

1. x_k^c, y_k^c acceptable to filter
2. $\|\nabla L(x_k, y_k, \rho_k) - z_k\| \leq \omega_k$
   ... optimality of Lagrangian

   New: $\omega_k := 0.1 \max \{\|\nabla L_l - z_l\|\}$
   ... depends on filter



Consequences of Filter

1. ensures that back-tracking line-search will succeed
   ... if not acceptable then reduce $\omega_{k+1} = \omega_k/2$
2. & $\omega_{k+1} = \omega_k/2$ ensure can always find Al Capone

A Filter for QPPAL

WHILE (not optimal) **BEGIN**

1. Find ω_k optimal solution x_k^c of

$$\underset{l \le x \le u}{\text{minimize}} \ \tfrac{1}{2} x^T H x + g^T x - y^T (A^T x - b) + \tfrac{1}{2} \rho_k \| A^T x - b \|^2$$

2. Find $\mathcal{A}(x_k^c)$ & estimate penalty $\bar{\rho} = 2 \, \| H_{\mathcal{I}} \| / \| A_{\mathcal{I}} A_{\mathcal{I}}^T \|$
3. **IF** $\bar{\rho} > \rho_k$ **THEN** update $\rho_{k+1} = \bar{\rho}$ & **CYCLE**
   ELSE update multiplier: $y_k^c = y_k - \rho_k (A^T x_k^c - b)$

5. Solve equality QP in subspace $\rightarrow \ (\Delta x_{\mathcal{I}}, \Delta y)$

END

A Filter for QPPAL

WHILE (not optimal) **BEGIN**

1. Find ω_k optimal solution x_k^c of

$$\underset{l \le x \le u}{\text{minimize}} \ \frac{1}{2}x^T H x + g^T x - y^T(A^T x - b) + \frac{1}{2}\rho_k \|A^T x - b\|^2$$

2. Find $\mathcal{A}(x_k^c)$ & estimate penalty $\bar{\rho} = 2\,\|H_{\mathcal{I}}\|/\|A_{\mathcal{I}} A_{\mathcal{I}}^T\|$
3. **IF** $\bar{\rho} > \rho_k$ **THEN** update $\rho_{k+1} = \bar{\rho}$ & **CYCLE**
   ELSE update multiplier: $y_k^c = y_k - \rho_k(A^T x_k^c - b)$
4. **IF** (x_k^c, y_k^c) not acceptable **THEN** $\omega_{k+1} = \omega_k/2$ & **CYCLE**
5. Solve equality QP in subspace $\to$ $(\Delta x_{\mathcal{I}}, \Delta y)$
6. Filter-search along $(x_k^c + \alpha\Delta x, y_k^c + \alpha\Delta y)$; **update** x, y, k

END

Properties of QPPAL

Lemma (Friedlander)

First-order multiplier update & augmented system solve $\Leftrightarrow$ *Newton step on first order conditions* $(\nabla_x \mathcal{L}, \nabla_y \mathcal{L}) = 0$

Properties of QPPAL

Lemma (Friedlander)

First-order multiplier update & augmented system solve $\Leftrightarrow$ *Newton step on first order conditions* $(\nabla_x \mathcal{L}, \nabla_y \mathcal{L}) = 0$

Filter version has no restoration phase

- Step 4. **IF** (x_k^c, y_k^c) not acceptable **THEN**

  $\omega_{k+1} = \omega_k / 2$ & **CYCLE**

  $\Rightarrow$ tighten tolerance of BCL $\Rightarrow$ BCL is restoration phase

  $\Rightarrow$ always find Al Capone (Cauchy point)

Properties of QPPAL

Lemma (Friedlander)

First-order multiplier update & augmented system solve $\Leftrightarrow$ *Newton step on first order conditions* $(\nabla_x \mathcal{L}, \nabla_y \mathcal{L}) = 0$

Filter version has no restoration phase

- Step 4. **IF** (x_k^c, y_k^c) not acceptable **THEN**
  $\omega_{k+1} = \omega_k/2$ & **CYCLE**
  $\Rightarrow$ tighten tolerance of BCL $\Rightarrow$ BCL is restoration phase
  $\Rightarrow$ always find Al Capone (Cauchy point)
- Augmented system solve is 2nd restoration phase:
  $A_{:,\mathcal{I}} x_{\mathcal{I}} = b - A_{:,\mathcal{A}} x_{\mathcal{A}}$ & full step $\Rightarrow$ feasibility

Outline

Two-Phase Active-Set Framework for NLP

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

repeat

1. Compute cheap first-order step $x^{(k)} + s$, e.g. LP/QP solve
2. Predict active set from s: $\mathcal{A}(x^{(k)} + s)$ & $\mathcal{I}(x^{(k)} + s)$
3. Compute second-order *EQP step* on active set:

$$\begin{bmatrix} H_k & A_k \\ A_k^T & \end{bmatrix} \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} = \ldots \qquad \text{Newton step}$$

   where $H_k = \nabla^2 L^{(k)}$ and $A_k = \left[\nabla c^{(k)} : I^{(k)}\right]$ active c/s

4. Enforce global convergence & set $k \leftarrow k + 1$

until optimal solution found

(Fletcher & de la Maza:89), (Gould & Robinson:10), (Fletcher:11)

Toward scalable nonlinear optimization
$\Rightarrow$ replace LP/QP ... avoid pivoting, i.e. rank-one matrix updates

Augmented Lagrangian Methods (LANCELOT)

Augmented Lagrangian:

$$L_\rho := f(x) - y^T c(x) + \frac{\rho}{2}\|c(x)\|_2^2$$

With sequences $\omega_k \searrow 0$ and $\eta_k \searrow 0$

repeat

1. Find ω_k optimal solution $\hat{x}^{(k+1)}$ of $\underset{x \geq 0}{\text{minimize}} \; L_\rho(x, y^{(k)})$
2. **if** $\|c(\hat{x}^{(k+1)})\| \leq \eta_k$ **then**

   update multipliers: $y^{(k+1)} = y^{(k)} - \rho_k c(\hat{x}^{(k+1)})$

   else

   increase penalty: $\rho_{k+1} = 2\rho_k$
3. Choose new $(\eta_{k+1}, \omega_{k+1})$; set $k \leftarrow k + 1$

until (optimal solution found)

see e.g. (Conn, Gould & Toint:95) and (Friedlander, 2002)

Augmented Lagrangian Methods (LANCELOT)

Advantage of Augmented Lagrangian Methods

- Scalable computational kernels

Disadvantages of Augmented Lagrangian Methods

1. First-order method in multipliers $\Rightarrow$ slow convergence
2. Arbitrary forcing sequences (ω_k, η_k) ... one fits all NLPs?
3. Slow penalty update $\Rightarrow$ slow for infeasible NLPs

Improving augmented Lagrangian methods:

1. Add equality QP step for fast Newton-like convergence
2. Replace forcing sequence (ω_k, η_k) by filter
3. Exploit structure for penalty estimates & use restoration phase

Goal: extend (Friedlander & L., 2008) from QP to NLP

Augmented Lagrangian Filter

Filter $\mathcal{F}$ to replace forcing sequences (ω_k, η_k)

Definition (Augmented Lagrangian Filter)

- Filter $\mathcal{F}$ is a list of pairs $(\eta(x), \omega(x,y))$ where

$$\omega(x,y) := \| \min\{x, \nabla_x L_0(x,y)\} \| \quad \text{... Lagrangian } L_0$$

$$\eta(x) := \|c(x)\| \quad \text{... constraint violation}$$

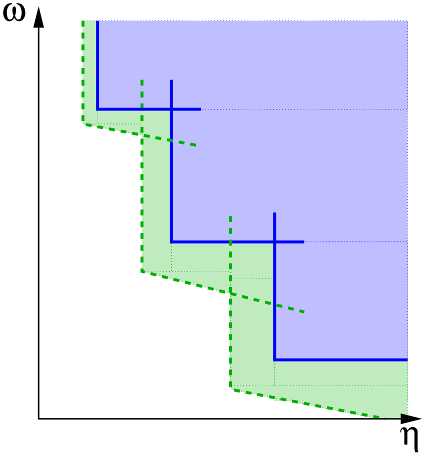
  such that no pair dominates another
- A point $(x^{(k)}, y^{(k)})$ acceptable to filter $\mathcal{F}$ iff

$$\eta(x^{(k)}) \leq \beta \eta_l \quad \text{or} \quad \omega(x^{(k)}, y^{(k)}) \leq \beta \omega_l - \gamma \eta(x^{(k)}), \quad \forall l \in \mathcal{F}$$

Typically: $\beta = 0.99$, $\gamma = 0.01$

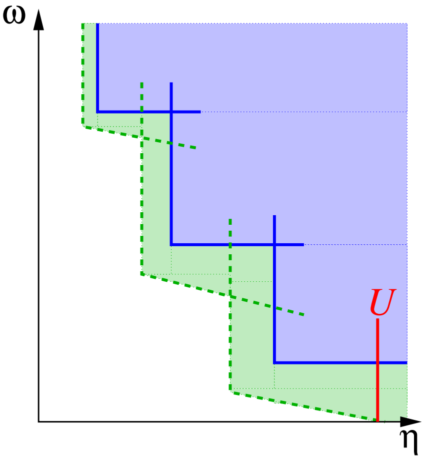
Approximate minimization of $L_\rho(x, y^{(k)})$ until acceptable to filter

Augmented Lagrangian Filter



- $\omega(x,y) := \| \min\{x, \nabla_x L_0(x,y)\}\|$ and $\eta(x) := \|c(x)\|$

Augmented Lagrangian Filter



- $\omega(x, y) := \| \min\{x, \nabla_x L_0(x, y)\} \|$ and $\eta(x) := \|c(x)\|$
- Automatic upper bound: $U = \beta / \gamma \omega_{\max}$, because $\omega \geq 0$

Augmented Lagrangian Filter Method

while $(x^{(k)}, y^{(k)})$ *not optimal* **do**

  $j = 0$; initialize $\hat{x}^{(j)} = x^{(k)}$, $\hat{\omega}_j = \omega_k$ and $\hat{\eta}_j = \eta_k$

  repeat

    $\hat{x}^{(j+1)} \leftarrow$ approximate $\text{argmin}_{x \geq 0} L_{\rho_k}(x, y^{(k)})$ from $\hat{x}^{(j)}$

    if *restoration switching condition* **then**

      Increase penalty: $\rho_{k+1} = 2\rho_k$ & switch to restoration

      ... find acceptable $(x^{(k+1)}, y^{(k+1)})$ and set $k = k + 1$

    end

    Provisionally update: $\hat{y}^{(j+1)} = y^{(k)} - \rho_j c(\hat{x}^{(j+1)})$

    Compute $(\hat{\eta}_{j+1}, \hat{\omega}_{j+1})$ and set $j = j + 1$

  until $(\hat{\eta}_j, \hat{\omega}_j)$ *acceptable to* $\mathcal{F}_k$;

  Set $(x^{(k+1)}, y^{(k+1)}) = (\hat{x}^{(j)}, \hat{y}^{(j)})$

  Get $\mathcal{A}^{(k+1)} = \{i : x_i^{(k+1)} = 0\}$ & solve equality QP

  if $\eta_{k+1} > 0$ **then** add $(\eta_{k+1}, \omega_{k+1})$ to $\mathcal{F}$ ... set $k = k + 1$;

end

Approximate Minimization of Augmented Lagrangian

Inner initialization: $j = 0$ and $\hat{x}^{(0)} = x^{(k)}$

For $j = 0, 1, \ldots$ terminate augmented Lagrangian minimization,

$$\hat{x}^{(j+1)} \leftarrow \text{approximate} \underset{x \geq 0}{\text{argmin}} \; L_{\rho_k}(x, y^{(k)})$$

when standard sufficient reduction holds:

$$\Delta L_{\rho_k} := L_{\rho_k}(\hat{x}^{(j)}, y^{(k)}) - L_{\rho_k}(\hat{x}^{(j+1)}, y^{(k)}) \geq \sigma \hat{\omega}_j \geq 0$$

E.g. Cauchy step on augmented Lagrangian for fixed ρ_k and $y^{(k)}$

More natural than requiring reduction in F.O. error $\hat{\omega}_j \searrow 0$

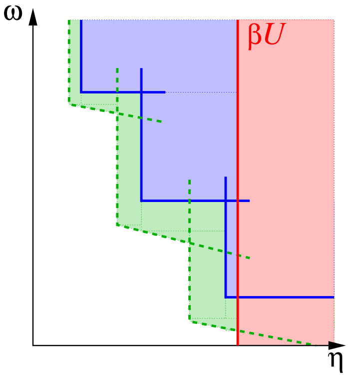
Switching to Restoration

Goal: Infeasible NLPs $\Rightarrow$ want to find $\underset{x \geq 0}{\text{minimize}} \; \|c(x)\|_2^2$ fast!

Switch to restoration, min $\|c(x)\|$, if

1. $\hat{\eta}_j \geq \beta U$ ... infeasible, or
2. $\hat{\eta}_j \geq M \min(1, \hat{\omega}_j^\tau)$, for $\tau \in [1, 2]$ ... infeasible Fritz-John point, or
3. $\|\min(\nabla c^{(j)} c^{(j)}, \hat{x}^{(j)})\| \leq \epsilon$ and $\|c^{(j)}\| \geq \beta \eta_{\min}$ ... infeasible FO Point, where

$$\eta_{\min} := \min_{l \in \mathcal{F}_k} \{\eta_l\} > 0$$



Lemma (Finite Return From Restoration)

$\eta_l \geq \eta_{\min} \; \forall l \in \mathcal{F}_k \Rightarrow \exists \; x^{(k+1)}$ *acceptable or restoration converges*

Second-Order Steps & KKT Solves

- $x^{(k+1)} \leftarrow \underset{x \geq 0}{\text{minimize}} \;\; L_\rho(x, y^{(k)})$ ... predicts $\mathcal{A}(x^{(k+1)})$
- Accelerate convergence, by solving EQP with $\Delta x_{\mathcal{A}} = 0$:

$$\begin{bmatrix} \tilde{H}_{k+1} & \tilde{A}_{k+1} \\ \tilde{A}_{k+1}^T & \end{bmatrix} \begin{pmatrix} \Delta x_{\mathcal{I}} \\ \Delta y \end{pmatrix} = \begin{pmatrix} -\nabla f_{\mathcal{I}}^{(k+1)} \\ -c(x^{(k+1)}) \end{pmatrix}$$

  where $\tilde{H}_{k+1}$ is "reduced" Hessian wrt bounds ($\Delta x_{\mathcal{A}} = 0$)
- Line-search: $\alpha_{k+1} \in \{0\} \cup [\alpha_{\min}, 1]$ such that

$$(x^{(k+1)}, y^{(k+1)}) = (\hat{x}^{(k+1)}, \hat{y}^{(k+1)}) + \alpha_{k+1}(\Delta x^{(k+1)}, \Delta y^{(k+1)})$$

  $\mathcal{F}_k$-acceptable
  ... $\alpha_{k+1} = 0$ OK, because $(\hat{x}^{(k+1)}, \hat{y}^{(k+1)})$ was acceptable

Augmented Lagrangian Filter Method

while $(x^{(k)}, y^{(k)})$ *not optimal* **do**

  $j = 0$; initialize $\hat{x}^{(j)} = x^{(k)}$, $\hat{\omega}_j = \omega_k$ and $\hat{\eta}_j = \eta_k$

  repeat

    $\hat{x}^{(j+1)} \leftarrow$ approximate $\text{argmin}_{x \geq 0} L_{\rho_k}(x, y^{(k)})$ from $\hat{x}^{(j)}$

    if *restoration switching condition* **then**

      Increase penalty: $\rho_{k+1} = 2\rho_k$ & switch to restoration

      ... find acceptable $(x^{(k+1)}, y^{(k+1)})$ and set $k = k + 1$

    end

    Provisionally update: $\hat{y}^{(j+1)} = y^{(k)} - \rho_j c(\hat{x}^{(j+1)})$

    Compute $(\hat{\eta}_{j+1}, \hat{\omega}_{j+1})$ and set $j = j + 1$

  until $(\hat{\eta}_j, \hat{\omega}_j)$ *acceptable to* $\mathcal{F}_k$;

  Set $(x^{(k+1)}, y^{(k+1)}) = (\hat{x}^{(j)}, \hat{y}^{(j)})$

  Get $\mathcal{A}^{(k+1)} = \{i : x_i^{(k+1)} = 0\}$ & solve equality QP

  if $\eta_{k+1} > 0$ **then** add $(\eta_{k+1}, \omega_{k+1})$ to $\mathcal{F}$ ... set $k = k + 1$;

end

Overview of Convergence Proof

Assumptions

1. Functions $f(x)$ and $c(x)$ twice continuously differentiable
2. $\|c(x)\| \to \infty$ whenever $\|x\| \to \infty$ ... ignore EQP for analysis

Outline of Convergence Proof

1. Filter $\mathcal{F}_k \Rightarrow$ iterates, $x^{(k)}$ remain in compact set
2. Inner iteration is finite $\Rightarrow \exists$ convergent subsequence
3. Mechanism of filter $\Rightarrow$ limit points are feasible
4. Show limit points are stationary in two cases:
    1. Bounded penalty ... rely on filter
    2. Unbounded penalty ... classical augmented Lagrangian

Remark

Do not assume compactness, or bounded multipliers!

Iterates Remain in Compact Set

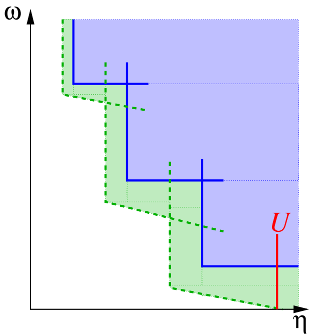
Lemma (All Iterates Remain in Compact Set)

All major and minor iterates, $x^{(k)}$ and $\hat{x}^{(j)}$ are in a compact set, C.

Proof.

1. Upper bound on filter ($U = \beta/\gamma\omega_{\max}$)
   $\Rightarrow \|c(x^{(k)})\| \leq U$ for all major iterates
2. Switching condition ($\hat{\eta}_j \leq \beta U$)
   $\Rightarrow \|c(\hat{x}^{(j)})\| \leq U$ for all minor iterates
3. Feasibility restoration minimizes $\|c(x)\|$
   $\Rightarrow \|c(x^{(k)})\|$ bounded

$\Rightarrow \|c(^{(k)})\| \leq U$ and $\|c(\hat{x}^{(j)})\| \leq U$



$c(x)$ twice continuously differentiable & $\|c(x)\| \to \infty$ if $\|x\| \to \infty$

$\Rightarrow x^{(k)}, \hat{x}^{(j)} \in C$, compact

Finiteness of Inner Iteration

Lemma (Finiteness of Inner Iteration)

The inner iteration is finite.

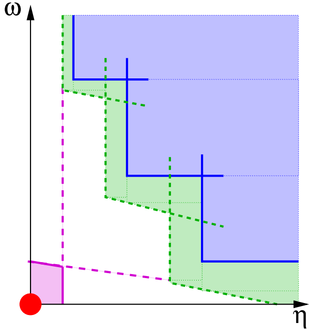
Proof. Assume inner iteration not finite $\Rightarrow \exists \hat{x}^* = \lim \hat{x}^{(j)} \in C$

1. Fixed penalty: $\rho_k \equiv \rho < \infty$
2. Sufficient reduction of $L_\rho(x, y^{(k)})$

   $\Rightarrow \Delta L_\rho \geq \sigma \hat{\omega}_j$; assume $\hat{\omega}_j \geq \bar{\omega} > 0$

   $\Rightarrow L_\rho(\hat{x}^{(j)}, y^{(k)})$ unbounded

   ... but $\|c(\hat{x}^{(j)})\|$, ρ, and $f(x)$ bounded
3. Contradiction $\Rightarrow \hat{\omega}_j \to 0$, and $\hat{\omega}_* = 0$
4. Switching: $\hat{\eta}_j < M\hat{\omega}_j \Rightarrow \hat{\eta}_* \leq M\hat{\omega}_*$

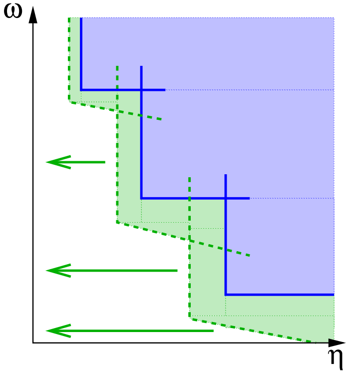


$\Rightarrow (\hat{\eta}_*, \hat{\omega}_*) = (0, 0)$ and $\exists$ filter acceptable points near $(0, 0)$

Feasible Limit Points

Lemma (Feasible Limit Points)

In outer iteration, feasibility error $\eta_k = \|c(x^{(k)})\| \to 0$.



Proof. Two cases:

1. $\eta_k = 0, \forall k \geq K_0$ ... of course!
2. $\eta_k > 0$, subsequence $\forall k \geq K_0$
   see (Chin & Fletcher, 2003)
   ... envelope $\Rightarrow \eta_k \to 0$

... standard filter argument

First-Order Optimality

Lemma (First-Order Stationarity)

First-order optimality $\omega_k = \| \min\{x^{(k)}, \nabla_x L_0^{(k)}\}\| \to 0.$

Proof. (1) $\rho_k \leq \bar{\rho} < \infty$ and (2) ρ_k unbounded: classical proof

- Assume $\omega_k \geq \bar{\omega} > 0$ & seek contradiction
  $\Rightarrow \Delta L_{\bar{\rho}}^{in} = L_{\bar{\rho}}(x^{(k)}, y^{(k)}) - L_{\bar{\rho}}(x^{(k+1)}, y^{(k)}) \geq \sigma \omega_k \geq \sigma \bar{\omega} > 0$
- First-order multiplier update, $y^{(k+1)} = y^{(k)} - \bar{\rho} c(x^{(k+1)})$

$$\begin{aligned} \Delta L_{\bar{\rho}}^{out} &= L_{\bar{\rho}}(x^{(k)}, y^{(k)}) - L_{\bar{\rho}}(x^{(k+1)}, y^{(k+1)}) \\ &= \Delta L_{\bar{\rho}}^{in} - \bar{\rho} \|c(x^{(k+1)})\|_2^2 \\ &\geq \sigma \bar{\omega} - \rho \|c(x^{(k+1)})\|_2^2 \end{aligned}$$

- Feasible limit: $c(x^{(k+1)}) \to 0 \Rightarrow \|c(x^{(k+1)})\|_2^2 \leq \sigma \frac{\bar{\omega}}{2\rho}, \; \forall k \geq \bar{K}$
  $\Rightarrow \Delta L_{\bar{\rho}}^{out} \geq \sigma \frac{\bar{\omega}}{2}, \; \forall k \geq \bar{K}$ outer iteration sufficient reduction

First-Order Optimality (Proof cont.)

- Sufficient reduction at outer iterations: $\Delta L_{\bar{\rho}}^{out} \geq \sigma \frac{\bar{\omega}}{2}$
  $\Rightarrow L_{\bar{\rho}}(x, y) = f(x) - y^T c(x) + \frac{\rho}{2}\|c(x)\|_2^2$ unbounded
- $x^{(k)} \in C$ compact $\Rightarrow f(x)$ and $\|c(x)\|_2^2$ bounded
- Show $y^T c(x) \leq M$ bounded:
  - Feasibility Lemma $\Rightarrow \eta_k = \|c(x^{(k)})\| \to 0$
  - Filter acceptance: Monotone sub-sequences $\eta_k \leq \beta \eta_{k-1}$
  - FO multiplier update: $y^{(k)} = y^{(0)} - \bar{\rho} \sum_l c^{(l)}$

$$\Rightarrow {y^{(k)}}^T c(x^{(k)}) = \left(y^{(0)} - \bar{\rho} \sum_l c^{(l)} \right)^T c^{(k)}$$

$$\leq \left(1 + \bar{\rho} \sum_l \eta_l \right) \eta_k \quad \leq \quad \eta_0 \left(\beta^k + \bar{\rho} \sum_l \beta^{l+k} \right) \quad \leq \quad M$$

- Contradiction: $L_{\bar{\rho}}(x, y) = f(x) - y^T c(x) + \frac{\rho}{2}\|c(x)\|_2^2$ bounded
  $\Rightarrow \omega_k \to 0$ ... first-order stationarity

Key Computational Kernels

1. Filter stopping rule readily included in minimization of L_ρ
   - $\nabla L_\rho(\hat{x}^{(j+1)}, y^{(k)}) = \nabla L_0(\hat{x}^{(j+1)}, \hat{y}^{(j+1)}) = \hat{\omega}_{j+1}$
2. Approximate minimization of augmented Lagrangian
   - projected gradient plus CG on subspace

$$\left[H_k + \rho A_k A_k^T\right]_{\mathcal{I},\mathcal{I}} \Delta x_{\mathcal{I}} = -\nabla L_\rho(x^{(k)}, y^{(k)})$$

$$\Leftrightarrow \begin{bmatrix} \tilde{H}_k & \tilde{A}_k \\ \tilde{A}_k^T & -\rho^{-1} I \end{bmatrix} \begin{pmatrix} \Delta x_{\mathcal{I}} \\ u \end{pmatrix} = \begin{pmatrix} -\nabla L_\rho(x^{(k)}, y^{(k)}) \\ 0 \end{pmatrix}$$

3. KKT system solve
   - $\begin{bmatrix} \tilde{H}_k & \tilde{A}_k \\ \tilde{A}_k^T & \end{bmatrix} \begin{pmatrix} \Delta x_{\mathcal{I}} \\ \Delta y \end{pmatrix} = \ldots$
   - indefinite reduced Hessian $\Rightarrow$ inertia control

$\Rightarrow$ exploit scalable matrix-free solvers based on H_k, A_k

Summary and Teaching Points

Presented Active-Set Method for QP/NLP

- Augmented Lagrangian Filter for QP or NLP
  - Identify active set via augmented Lagrangian step
  - Perform EQP solve
  - Use filter instead of forcing sequences
- Main computational kernels are parallelizable
  - Bound-constrained optimization via projected-gradient with CG
  - KKT-system solves, using `GMRES` or `MINRES`