An augmented Lagrangian trust region method for equality constrained optimization

Ya-xiang Yuan

State Key Laboratory of Scientific and Engineering Computing
Inst of Comput. Math. and Scientific/Engineering Computing
AMSS, Chinese Academy of Sciences, Beijing 100190, China
Email: yyx@lsec.cc.ac.cn
http://lsec.cc.ac.cn/∼yyx

Recent Advances on Optimization, July 24-26, 2013, Toulouse

Joint work with Xiao Wang

Outline

Outline

Outline

Outline

Outline

Outline

Introduction

General Equality Constrained Optimization:

$$\min_{x\in\mathbb{R}^n} \quad f(x) \tag{1a}$$

$$\text{s. t.} \quad c(x) = 0, \tag{1b}$$

where $c(x) = (c_1(x), \ldots, c_m(x))^T$

$f(x)$ and $c_i(x)$, $i = 1, \ldots, m$, are Lipschitz continuously differentiable.

Methods for Equality constrained Optimization

- Sequential Quadratic Programming(SQP)

$$\min_{d\in\mathbb{R}^n} \quad g_k^T d + \frac{1}{2} d^T B_k d \tag{2a}$$

$$\text{s. t.} \quad c_k + A_k d = 0, \tag{2b}$$

  Globalization techniques:

  - line search (Powell, Han, Gill, Murray, Wright, Schittkowski, ...)
  - trust region (Dennis, Tapia, Powell, Byrd, Nocedal, ...,)
  - filter (Fletcher, Leyffer, Toint, ...)

- Sequential l_1 Quadratic Programming (Sl_1QP) Method (Fletcher, 1982)
- Sequential l_∞ Quadratic Programming (Sl_∞QP) Method (Yuan, 1995)
- Penalty Function methods

Augmented Lagrangian method

For λ_k and σ_k, compute x_k by

$$\min_{x\in\mathbb{R}^n} \quad L(x;\lambda_k,\sigma_k) = f(x) - \lambda_k^T c(x) + \frac{\sigma_k}{2}\|c(x)\|_2^2. \tag{3}$$

Then update $\lambda_{k+1} = \lambda_k - \sigma_k c(x_k)$

- Dated back to
  Hestenes(1969), Powell(1969), Rockafellar(1973), ...
- Efficient Implementation: LANCELOT by
  Conn, Gould and Toint (Beale-Orchard-Hays prize winners)
- Applications
  - Uzawa Method for Linear Equations
  - Bregman iterative algorithms for Compressive Sensing
  - Low-rank matrix optimization

Algorithm of Niu and Yuan

L.F. Niu and Y. Yuan, “A new trust region algorithm for nonlinear constrained optimization”, *J. Comp. Math.* 28(2010) pp. 72-86.

Algorithm Features

- Trust region
- filter technique
- approximate Lagrangian

- Nice numerical results
- no theoretical results

Features of Our New Method

- Simple subproblem, similar to that of Niu and Yuan
- New strategy to update the penalty parameter σ_k
- A switching condition for updating the Lagrange multiplier
- Global convergence analysis

Quadratic Approximation to Augmented Lagrangian

Instead of

$$\min_{x \in \mathbb{R}^n} \quad L(x; \lambda_k, \sigma_k). \tag{4}$$

Consider

$$\begin{aligned}\Phi_k(d) &= L(x_k; \lambda_k, \sigma_k) + \nabla_x l(x_k; \lambda_k)^T d + \frac{1}{2} d^T B_k d + \frac{\sigma_k}{2} \|c_k + A_k d\|^2 \\ &= L(x_k; \lambda_k, \sigma_k) + g_k^T d - \lambda_k^T A_k d + \frac{1}{2} d^T B_k d + \frac{\sigma_k}{2} \|c_k + A_k d\|^2,\end{aligned} \tag{5}$$

where $B_k = (\approx) \nabla_x^2 l(x_k, \lambda_k)$

New Subproblem

$$\min_{d\in\mathbb{R}^n} \quad q_k(d) = g_k^T d - \lambda_k^T A_k d + \frac{1}{2} d^T B_k d + \frac{\sigma_k}{2} \|c_k + A_k d\|^2 \quad \text{(6a)}$$

$$\text{s. t.} \quad \|d\| \le \Delta_k. \quad \text{(6b)}$$

B_k can be updated by quasi-Newton updates

Inexact Solutions of the Subproblem

Inexact solution s_k that satisfies

$$q_k(0) - q_k(s_k) \geq \bar{\beta} \cdot [q_k(0) - \min_{\|d\| \leq \Delta_k} q_k(d)] \tag{7}$$

for some positive constant $\bar{\beta} \in (0, 1)$.

- Truncated CG step (Toint-Steihaug algorithm) satisfies (7) with $\bar{\beta} = 0.5$ for convex subproblems (Yuan, 2000)
- Assume that (7) holds for all k

Update of Trust Region Bound Δ_k

After obtaining s_k, Compute

$$\rho_k = \frac{\text{Ared}_k}{\text{Pred}_k} = \frac{L(x_k;\lambda_k,\sigma_k) - L(x_k + s_k;\lambda_k,\sigma_k)}{q_k(0) - q_k(s_k)} \tag{8}$$

General Trust Region Properties

- $\rho_k \to 1$ as $\Delta_k \to 0$. (at a nonstationary point)
- good iteration if $\rho_k \geq \eta$

Special situation:

$$\nabla_x L(x_k;\lambda_k,\sigma_k) = 0 \quad \text{and} \quad \|c_k\| > 0. \tag{9}$$

Update of Penalty parameter

Updating technique of Niu and Yuan

$$\sigma_{k+1} = \begin{cases} \max\{2\sigma_k, 2\|\lambda_{k+1}\|\}, & \text{if } \|c(x_k + s_k)\| \ \geq \ 0.5\|c_k\|, \\ \max\{\sigma_k, 2\|\lambda_{k+1}\|\}, & \text{otherwise.} \end{cases} \tag{10}$$

- The motivation for (10) is the inequality $\sigma_{k+1} \geq 2\|\lambda_{k+1}\|$
- However, $\sigma_* \geq 2\|\lambda_*\|$ is not a necessary condition for

$$L(x; \lambda_*, \sigma_*) \ = \ f(x) - \lambda_*^T c(x) + \frac{\sigma_*}{2}\|c(x)\|^2$$

to be an exact penalty function of (1)

Updating σ_k

General Approach σ_k will be increased if

$$\|c(x_k + s_k)\| \geq \tau \|c_k\|$$

for some constant $0 < \tau < 1$

Our Approach

- penalty parameter σ_k is increased if

$$\text{Pred}_k < \delta_k \sigma_k \min\{\Delta_k \|c_k\|, \|c_k\|^2\}, \quad (11)$$

- $\sigma_k \delta_k$ to have the property

$$\sigma_k \delta_k \to 0 \quad (12)$$

if $\sigma_k \to \infty$

Update of the Multipliers: When to update

switch condition

$$\|c_{k+1}\| \leq R_j, \tag{13}$$

$\{R_j\}$ is a sequence of nonincreasing controlling factors.

Update λ only when the switch condition holds

Update of Multipliers: How to update

As we do not minimize the Augmented Lagrangian, we do not have

$$g_{k+1} - A_{k+1}^T \lambda_k + \sigma_k A_{k+1}^T c_{k+1} = 0, \tag{14}$$

Thus, it might not be reasonable to use $\lambda_{k+1} = \lambda_k - \sigma_k c_{k+1}$.

We use

$$\tilde{\lambda}_k = arg \min_{\lambda \in \mathbb{R}^m} \|g_k - A_k^T \lambda\|, \tag{15}$$

with

$$\lambda_k = P_{[\lambda_{min}, \lambda_{max}]} \tilde{\lambda}_k; \tag{16}$$

Augmented Lagrangian Trust Region Method (ALTR)

Step 0 **Initialization.** Given constants $\beta \in (0,1)$, $\epsilon > 0$, $\theta > 1$,
$\lambda_{\min} < \lambda_{\max}$ and $0 < \eta < \eta_1 < \frac{1}{2}$, $\quad R_0 = \max\{\|c(x_0)\|, 1\}$
Given $x_0 \in \mathbb{R}^n$, $B_0 \in \mathbb{R}^{n\times n}$, $\lambda_0 \in \mathbb{R}^m$, $\sigma_0 > 0$, $\delta_0 > 0$, $\Delta_0 > 0$;
Set $j := 0$, $k := 0$.

Step 1 **Termination Test.**
If $\|c_k\| = 0$ and $P_{\mathcal{N}_k}(g_k) = 0$, stop (return x_k as solution).

Step 2 **Computing Trial Step.**
While $\nabla_x L(x_k; \lambda_k, \sigma_k) = 0$ and $\|c_k\| > 0$,

$$\sigma_k := \theta\sigma_k; \tag{17}$$

Endwhile;
Solve (6) obtaining s_k;

Step 3 **Update of Iterates.**
Compute the ratio ρ_k in (8);
If $\rho_k \geq \eta$, go to Step 4;
$\Delta_{k+1} = \|s_k\|/4$; $x_{k+1} = x_k$; $k := k+1$; go to Step 2;

Step 4 **Update of Penalty Parameter and Multiplier.**

If $\text{Pred}_k < \delta_k \sigma_k \min\{\Delta_k \|c_k\|, \|c_k\|^2\}$, then
$\sigma_{k+1} = 2\sigma_k; \quad \delta_{k+1} = \delta_k/4$; else $\sigma_{k+1} = \sigma_k$, $\delta_{k+1} = \delta_k$;
If $\|c_{k+1}\| \le R_j$, then compute

$$\begin{aligned} \tilde{\lambda}_{k+1} &= arg\min \|A_{k+1}^T \lambda - g_{k+1}\|; \\ \lambda_{k+1} &= P_{[\lambda_{\min}, \lambda_{\max}]} \tilde{\lambda}_{k+1}; \qquad R_{j+1} := \beta R_j; \end{aligned}$$

else $\lambda_{k+1} = \lambda_k$;

Step 5 **Update of Trust Region Radius.**

Set $x_{k+1} = x_k + s_k$;

$$\Delta_{k+1} = \begin{cases} \max\{\Delta_k, 1.5\|s_k\|\}, & \text{if } \rho_k \in [1-\eta_1, +\infty), \\ \Delta_k, & \text{if } \rho_k \in [\eta_1, 1-\eta_1), \\ \max\{0.5\Delta_k, 0.75\|s_k\|\}, & \text{if } \rho_k \in [\eta, \eta_1); \end{cases}$$

Calculate f_{k+1}, g_{k+1}, c_{k+1} and A_{k+1}; Generate B_{k+1};
$k := k+1$, $j := j+1$, then go to Step 1.

Convergence: Assumptions

We assume

AS.1 $f(x)$ *and* $c(x)$ *are Lipschitz continuously differentiable.*

AS.2 $\{x_k\}$ *and* $\{B_k\}$ *are uniformly bounded.*

Global Convergence: Lemmas

Lemma

*Under assumptions **AS.1**-**AS.2**, if* $\sigma_k \to \infty$, *then* $\lim_{k\to\infty} \|c_k\|$ *exists.*

Idea of Proof

- the monotone property of σ_k,
- Analyzing $\sum \sigma_k^{-1}[L(x_k, \lambda_k, \sigma_k) - L(x_{k+1}, \lambda_k, \sigma_k)]$
- $||c(x_{q+1})||^2 \leq ||c(x_q)||^2 + 2M_0\sigma_p^{-1}$
- $||c(x_{q+1}||^2 < \liminf ||c(x_k)||^2 + \epsilon$

due to Professor Powell

Convergence Theory: Infeasible Case

Theorem

Under assumptions ***AS.1-AS.2****, if* $\lim_{k\to\infty} \sigma_k = \infty$ *and* $\lim_{k\to\infty} \|c_k\| > 0$*, then any accumulation point of* $\{x_k\}$ *is an infeasible stationary point of the least square problem*

$$\min_{x\in\mathbb{R}^n} \quad \|c(x)\|^2. \tag{18}$$

Convergence Theory: Feasible Case

Theorem

Under assumptions ***AS.1-AS.2****, if* $\lim_{k\to\infty} \sigma_k = \infty$ *and* $\lim_{k\to\infty} \|c_k\| = 0$*, then the sequence of iterates* $\{x_k\}$ *is not bounded away from KKT points, or Fritz-John points at which CPLD fails to hold.*

Proof

Let $\bar{x}$ be any accumulation point of $\{x_k\}$, if $\lim_{k\to\infty} \|c_k\| = 0$ and CPLD fails to hold at $\bar{x}$, $\implies$ $\bar{x}$ is a Fritz-John point.

We next study the case when CPLD holds:
Two possible cases may happen.

Case 1. **Update** (17) **occurs only finitely many times.**

Without loss of generality, we assume (17) never happens.
Firstly, we prove that for any $\varepsilon > 0$, there exists $k = k(\varepsilon)$, such that

$$\|c_k\| < \varepsilon \quad \text{and} \quad \|P_{\mathcal{N}_k}(g_k - A_k^T\lambda_k)\| < \varepsilon. \tag{19}$$

The above inequality is proved by contradiction, using

$$\begin{aligned} \text{Pred}_k &\geq \bar{\beta}\xi_1\|P_{\mathcal{N}_k}[g_k - A_k^T\lambda_k]\|\min\{\xi_2\|P_{\mathcal{N}_k}[g_k - A_k^T\lambda_k]\|, \Delta_k\} \\ &\geq \bar{\beta}\xi_1\bar{\varepsilon}\min\{\xi_2\bar{\varepsilon}, \Delta_k\} \\ &\geq \bar{\nu}\min\{\Delta_k\|c_k\|, \|c_k\|^2\}, \end{aligned}$$

Consequently, by forcing $\varepsilon \to 0$ we obtain a subsequence $\mathcal{K}$ such that

$$\lim_{\substack{k\to\infty \\ k\in\mathcal{K}}} \|c_k\| = 0 \quad \text{and} \quad \lim_{\substack{k\to\infty \\ k\in\mathcal{K}}} \|P_{\mathcal{N}_k}(g_k - A_k^T \lambda_k)\| = 0, \tag{20}$$

Assume x_* is any accumulation point of $\{x_k\}_{\mathcal{K}}$ at which CPLD holds. We next prove that x_* is a KKT point.

Denote $\{\nabla c_i(x_*)\}_{i\in I}$ as the maximal set of linearly independent vectors, among all the gradients of constraints at x_*.

Then there is an index k_0 such that for all $k_0 < k \in \mathcal{K}$,

$$\{\nabla c_i(x_k)\}_{i\in I} \quad \text{and} \quad \nabla c_j(x_k), \quad j \notin I$$

are linearly dependent.

It implies that

$$\text{Range}(A_k^T) = \text{span}\{\nabla c_i(x_k), i = 1, \ldots, m\} = \text{span}\{\nabla c_i(x_k), i \in I\}$$

Define

$$\bar{A}_k^T = (\nabla c_i(x_k))_{i \in I}. \tag{21}$$

and $\bar{A}_*^T = (\nabla c_i(x_*))_{i \in I}$.

There exist k_0 and M such that

$$\|(\bar{A}_k^T)^+\| \le M, \quad \forall k \in \mathcal{K}_1, k \ge k_0.$$

Let $y_k = (\bar{A}_k^T)^+ g_k$, which gives

$$g_k - \mu_k = \bar{A}_k^T y_k, \quad \forall k \in \mathcal{K}_1, k \ge k_0.$$

$\Longrightarrow$

$$g_* = \bar{A}_*^T y_*.$$

$\Longrightarrow$ $g_* \in \text{Span}\{\nabla c_i(x_*), i \in \mathcal{I}\}$, $\quad \Longrightarrow$ x_* is a KKT point.

Case 2. **Update** (17) **occurs in infinitely many iterations.**

In this case, we can show that there exist a subsequence of $\{x_k\}_{\mathcal{K}}$ and $\{\bar{\sigma}_k\}_{\mathcal{K}}$ such that

$$g_k - A_k^T \lambda_k + \bar{\sigma}_k A_k^T c_k = 0, \quad k \in \mathcal{K}. \tag{22}$$

Suppose $\{x_k\}_{\mathcal{K}_1} \to x_*$ where $\mathcal{K}_1 \subseteq \mathcal{K}$. If CPLD is satisfied at $x = x_*$, for all sufficiently large $k \in \mathcal{K}_1$. there exist $\bar{A}_k$ which has full row rank such that

$$\text{Range}(A_k^T) = \text{Range}(\bar{A}_k^T). \tag{23}$$

Therefore,

$$g_* \in \text{Range}(\bar{A}_*^T)$$

due to $\{x_k\}_{\mathcal{K}_1} \to x_*$. It follows from $c_* = 0$ that x_* is a KKT point.

Convergence Theory: Bounded σ_k

Lemma

*Under assumptions **AS.1**-**AS.2**, if $\{\sigma_k\}$ is bounded, then there must have*

$$\lim_{k\to\infty} \|c_k\| = 0. \tag{24}$$

That is to say, all accumulation points of $\{x_k\}$ are feasible.

Theorem

*Under assumptions **AS.1**-**AS.2**, if $\{\sigma_k\}$ is bounded from above, then $\{x_k\}$ generated by Algorithm 2.1 is not bounded away from KKT points of (1).*

Proof of the Main Theory

Without loss of generality, assume $\sigma_k = \sigma$ for all k. Subproblem transformed into

$$\min \quad q_k(d) = \bar{g}_k^T d + \frac{1}{2} d^T (B_k + \sigma A_k^T A_k) d + \frac{\sigma}{2} \|c_k\|^2 \qquad (25a)$$

$$\text{s. t.} \quad \|d\| \le \Delta_k \qquad (25b)$$

with $\bar{g}_k = g_k - A_k^T \lambda_k + \sigma A_k^T c_k$.

$$\begin{aligned} q_k(0) - q_k(s_k) &\ge \bar{\beta} \frac{\|\bar{g}_k\|}{2} \min\{\frac{\|\bar{g}_k\|}{\|B_k + \sigma A_k^T A_k\|}, \Delta_k\} \\ &\ge \bar{\beta} \zeta_1 \|\bar{g}_k\| \min[\zeta_2 \|\bar{g}_k\|, \Delta_k], \qquad (26) \end{aligned}$$

with two positive constants ζ_1 and ζ_2.

Then,

$$\liminf_{k\to\infty} \|\bar{g}_k\| = 0. \tag{27}$$

Therefore, $\|c_k\| \to 0$, and

$$\liminf_{k\to\infty} \|g_k - A_k^T \lambda_k\| = 0.$$

As λ_k is bounded for all k, there exist an accumulation point x_* of $\{x_k\}$ and an accumulation point λ_* of $\{\lambda_k\}$ such that

$$c(x_*) = 0, \quad \nabla f(x_*) = A(x_*)^T \lambda_*.$$

Consequently, x_* is a KKT point.

Boundedness of the Penalty Parameters

AS.3 x_k *converges to* x_* *at which LICQ condition holds.*

Lemma

Under assumptions ***AS.1-AS.2****, assume that the sequence* $\{x_k\}$ *generated by Algorithm 2.1 converges to* x_* *at which* ***AS.3*** *holds. Then for all large* $k \in \mathcal{K}$,

$$\lambda_{min} < \tilde{\lambda}_k < \lambda_{max}. \quad (28)$$

Consequently, $\lambda_k = \tilde{\lambda}_k$ *for all large* $k \in \mathcal{K}$.

AS.4 $\lambda_k = (A_k^T)^+ g_k$ *holds for all large* k.

Theorem

Under assumptions ***AS.1-AS.4****, assume that the sequence* $\{s_k\}$ *is generated by Algorithm 2.1. If Algorithm 2.1 does not terminate finitely, then the predicted reduction obtained by* s_k *satisfies the inequality*

$$Pred_k \ \geq \ \delta_k \sigma_k \min\{\Delta_k \|c_k\|, \|c_k\|^2\} \tag{29}$$

for all sufficiently large k*. Consequently, all the penalty parameters* σ_k *are bounded from above.*

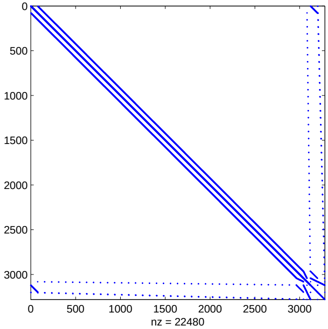
Numerical Experiments

- **Program:**
  Matlab 7.6.0 (R2008a)
  1.86 GHz Pentium Dual-Core microprocessor
  1GB of memory
  running Fedora 8.0.
- **Problems:**
  136 problems from CUTEr.
  Number of variables: from 2 to 4499
  Number of constraints: from 1 to 2998

Program Settings

- default starting point x_0 for each problem
- $B_k = \nabla^2_{xx} l(x_k, \lambda_k)$
- a modified verison of Moré and Sorensen's subroutine for the subproblem

Sparse Hessian



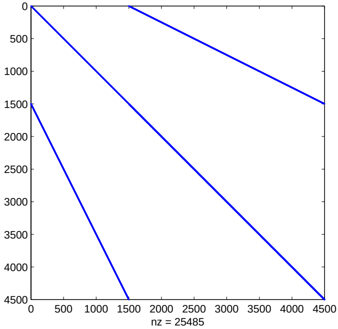


Figure: left: AUG2DC(3280×3280, nz: 0.2%); right: DTOC3(4499×4499, nz: 0.12%)

Parameters settings

- Initial trust region bound: $\Delta_0 = 1$.
- The termination condition

$$\|c_k\| < 10^{-5}, \quad \|P_{\mathcal{N}_k}(g_k)\| < 10^{-5}.$$

LANCELOT SETTINGS:

- gradient-accuracy-required 1e-5
- exact-second-derivatives-used
- trust-region-radius 1.0
- maximum-number-of-iterations 1000
- two-norm-trust-region-used

fmincon settings:

options = optimset('Algorithm', 'Active-set Algorithm', 'Hessian', 'on', 'InitTrustRegionRadius', '1', 'MaxFunEvals', '1000', 'TolCon', 1e-12, 'TolFun', '1e-6', 'TolPCG', 1e-5, 'TolX', 1e-15).

Results on Small and Medium Scale Problems – I

Problem	Prob. Dim.		LANCELOT		*fmincon*		ALTR	
	n	m	n_f	n_g	n_f	n_g	n_f	n_g
AIRCRFTA	8	5	4	5	3	2	3	3
ARGTRIG	200	200	19	17	3	2	4	4
ARTIF	102	100	34	25	12	10	11	11
BDVALUE	100	100	1	2	2	1	2	2
BDVALUES	100	100	28	29	14	13	33	33
BOOTH	2	2	3	4	2	1	4	4
BROWNALE	200	200	6	7	14	7	9	9
BROYDN3D	500	500	6	7	5	4	7	7
BT1	2	1	23	20	F		7	7
BT2	3	1	27	27	16	15	21	21
BT3	5	3	10	11	9	8	22	22
BT4	3	2	20	21	13	11	7	7
BT5	3	2	17	17	9	7	6	6
BT6	5	2	21	20	33	26	13	11
BT7	5	3	48	46	154	38	122	120
BT8	5	2	27	25	11	10	14	14
BT9	4	2	20	21	F		23	21
BT10	2	2	17	18	8	7	16	16
BT11	5	3	18	19	13	10	18	18
BT12	5	3	22	21	7	6	21	18
BYRDSPHR	3	2	35	22	166	14	16	15
CATENA	33	10	53	53	F	F	133	111
CATENARY	33	10	81	79	F	F	277	254
CHAIN	800	401	F		4	3	F	
CHANDHEU	100	100	14	15	10	9	13	13
CHNRSBNE	50	98	74	61	F	F	46	40

Results on Small and Medium Scale Problems – 2

Problem	Prob. Dim.		LANCELOT		*fmincon*		ALTR	
	n	m	n_f	n_g	n_f	n_g	n_f	n_g
CLUSTER	2	2	11	11	8	7	9	9
CUBENE	2	2	46	40	3	2	19	15
DECONVNE	61	40	57	46	2	1	24	15
DRCAVTY3	196	100	45	37	F		20	12
DTOC2	298	198	31	31	114	52	82	71
EIGENA2	6	3	5	6	3	2	5	5
EIGENACO	110	55	19	20	3	1	13	13
EIGENAU	110	110	20	20	2	1	12	12
EIGENB2	6	3	10	10	3	1	19	19
EIGENB	110	110	185	151	F		86	69
EIGENBCO	6	3	18	16	3	1	10	9
EIGENC2	30	15	44	40	2	1	10	10
EIGENCCO	462	231	204	169	F		211	198
ELEC	75	25	48	42	667	211	27	22
FLOSP2TH	323	323	F				230	158
GENHS28	10	8	6	7	8	7	8	8
GOTTFR	2	2	27	24	8	5	10	6
HATFLDF	3	3	24	22	F		9	7
HATFLDG	25	25	15	14	18	6	8	8
HEART6	6	6	F		F	F	490	481
HEART8	8	8	599	520	F	F	40	35
HIMMELBA	2	2	3	4	2	1	5	5
HIMMELBC	2	2	7	7	7	5	6	6
HIMMELBE	3	3	5	6	3	2	5	5
HS100LNP	7	2	39	38	F		9	7
HS111LNP	10	3	55	52	42	41	12	12

Results on Small and Medium Scale Problems – 3

Problem	Prob. Dim.		LANCELOT		*fmincon*		ALTR	
	n	m	n_f	n_g	n_f	n_g	n_f	n_g
HS26	3	1	31	29	8	4	16	16
HS27	3	1	12	13	631	96	13	11
HS28	3	1	3	4	8	6	6	6
HS39	4	2	20	21	F		23	21
HS40	4	3	10	11	7	6	6	6
HS42	4	2	10	11	10	9	8	8
HS46	5	2	26	20	15	11	17	16
HS47	5	3	22	22	61	27	16	15
HS48	5	2	3	4	8	6	6	6
HS49	5	2	15	16	22	18	15	15
HS50	5	3	10	11	19	10	14	14
HS51	5	3	2	3	7	5	10	10
HS52	5	3	7	8	6	5	17	17
HS56	7	4	13	13	10	12	10	8
HS61	3	2	18	18	F		11	10
HS6	2	1	53	48	14	7	14	12
HS77	5	2	23	22	24	21	13	10
HS78	5	3	13	12	10	9	7	7
HS79	5	3	14	13	11	10	6	6
HS7	2	1	17	17	18	10	8	8
HS8	2	2	10	10	6	5	6	6
HS9	2	1	5	6	11	6	6	6
HYDCAR20	99	99	F		10	8	759	753
HYDCAR6	29	29	F		6	5	76	71
HYPCIR	2	2	5	6	6	4	5	5
INTEGREQ	102	100	3	4	3	2	3	3

Results on Small and Medium Scale Problems – 4

Problem	Prob. Dim.		LANCELOT		*fmincon*		ALTR	
	n	m	n_f	n_g	n_f	n_g	n_f	n_g
JUNKTURN	510	350	72	68	3	2	F	
LCH	150	1	35	34	F		29	17
MARATOS	2	1	7	8	4	3	8	6
MWRIGHT	5	3	18	18	19	9	8	8
METHANB8	31	31	243	244	3	2	4	4
METHANL8	31	31	592	584	5	4	18	18
MSQRTA	100	100	18	17	7	5	12	11
MSQRTB	100	100	19	17	7	4	9	9
OPTCTRL3	299	200	82	83	F		31	28
OPTCTRL6	299	200	82	83	F		28	26
ORTHRDM2	203	100	300	260	8	6	8	8
ORTHRDS2	503	250	852	738	F		614	614
ORTHREGB	27	6	74	66	6	5	48	47
ORTHRGDS	503	250	908	790	43	32	F	
POWELLBS	2	2	47	42	23	11	61	56
POWELLSQ	2	2	16	14	F		22	19
RECIPE	3	3	16	17	23	11	14	14
RSNBRNE	2	2	32	28	5	2	14	12
S316-322	2	1	26	27	F		14	14
SINVALNE	2	2	37	33	5	2	21	18
SPMSQRT	499	829	15	13	F		10	10
TRIGGER	7	6	22	20	29	11	12	10
YATP1SQ	120	120	181	161	11	5	77	70
YATP2SQ	120	120	993	905	F		20	20
YFITNE	3	17	95	81	F		39	38
ZANGWIL3	3	3	7	8	3	1	11	11

Results on Large Scale Problems – 1

Problem	Prob. Dim. n	Prob. Dim. m	LANCELOT nf	LANCELOT ng	LANCELOT CPU(s)	*fmincon* nf	*fmincon* ng	*fmincon* CPU(s)	ALTR nf	ALTR ng	ALTR CPU(s)
AUG2DC	3280	1600	58	59	1.99	3	1	462.65	29	29	3.77
BRATU2D	1024	900	4	5	0.11			$>$10m	7	7	0.44
BRATU2DT	1024	900	8	9	0.26			$>$10m	9	9	0.56
BROYDN3D	1000	1000	6	7	0.02	9	4	20.964	8	8	0.23
CBRATU2D	3200	2888	5	6	0.55			$>$10m	8	8	1.6
CBRATU3D	3456	2000	6	7	0.15			$>$10m	7	7	3.14
DRCAVTY1	961	961	46	40	14.26	F			29	24	7.03
DRCAVTY2	961	961	104	85	24.06	F			64	59	16.5
DTOC1L	745	490	14	15	0.11	17	8	25.248	11	11	0.37
DTOC3	4499	2998	57	58	0.87			F	30	30	1.92
DTOC4	4499	2998	33	34	0.84			F	28	28	1.52
DTOC5	1999	999	37	38	0.3	38	11	385.4	26	26	0.52
DTOC6	1000	500	120	118	0.98			$>$10m	128	197	1.46
EIGENC	462	462	299	247	9.83	F			45	35	23.03
FLOSP2TL	867	803			$>$10m	9	4	336	21	21	6.6
FLOSP2TM	867	803			$>$10m	19	9	461.1	66	66	27.54
GRIDNETB	3444	1764	32	33	3.11			F	22	22	4.18

Results on Large Scale Problems – 2

	Prob. Dim.		LANCELOT			*fmincon*			ALTR		
Problem	*n*	*m*	*nf*	*ng*	CPU(s)	*nf*	*ng*	CPU(s)	*nf*	*ng*	CPU(s)
HAGER1	2001	1000	11	12	0.14			$>$10m	14	14	0.37
HAGER2	2001	1000	12	13	0.12	7	3	142.46	15	15	0.40
HAGER3	2001	1000	9	10	0.15	13	7	134.1	13	13	0.49
LUKVLE10	1000	998	47	37	0.2			$>$10m	24	19	0.47
LUKVLE11	998	664	34	30	0.14	21	4	18.884	26	26	0.43
LUKVLE13	998	664	101	93	0.29			$>$10m	121	116	1.56
LUKVLE16	997	747	59	50	0.17			$>$10m	42	36	0.86
LUKVLE1	1000	998	20	20	0.14	43	15	98.29	30	30	0.53
LUKVLE3	1000	2	26	26	0.08			$>$10m	16	16	0.27
LUKVLE6	999	499	41	42	0.31			$>$10m	20	20	0.53
LUKVLE7	1000	4	92	80	0.21			$>$10m	29	20	0.33
ORTHREGA	2053	1024	177	173	1.8			$>$10m	26	23	11.71
ORTHREGC	1005	500	50	44	0.27	46	21	91.14	17	11	1.61
ORTHREGD	1003	500	515	443	3.55			$>$10m	14	12	1.07
ORTHRGDM	4003	2000	158	141	2.96			F	12	12	27.19

Performance Profile (Dolan and Moré)

Solvers $\mathcal{S}$; Test set $\mathcal{P}$.
$t_{p,s}$ is the time that problem p solved by solver s. Define the performance ratio

$$r_{p,s} = \frac{t_{p,s}}{\min\{t_{p,s} : \quad s \in \mathcal{S}\}},$$

Set $r_{p,s} = +\infty$ if solver s does not solve problem p.
Performance Profile of solver s on the test set $\mathcal{P}$:

$$\rho_s(\tau) = \frac{1}{|\mathcal{P}|}\text{size}\{p \in \mathcal{P} : \log_2 r_{p,s} \leq \tau\}.$$

- $\rho_s(\tau)$: probability that $r_{p,s}$ is within the factor 2^τ.
- $\rho_s(0)$ is the probability that the solver s wins over the rest of solvers.

Compare with LANCELOT

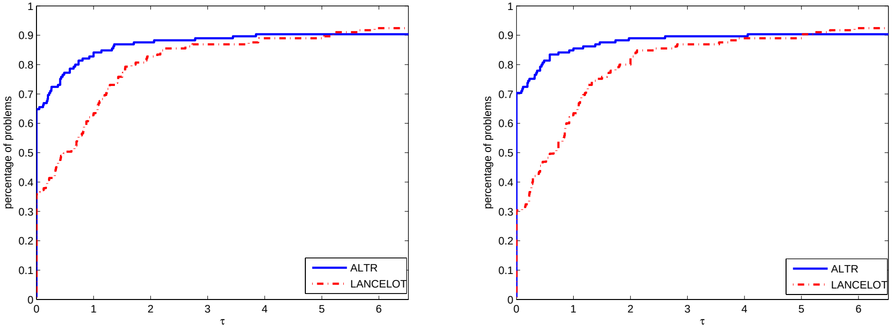


Figure: Comparison between ALTR and LANCELOT on all problems: function evaluations (left) and gradient evaluations (right)

Compare with matlab function fmincon

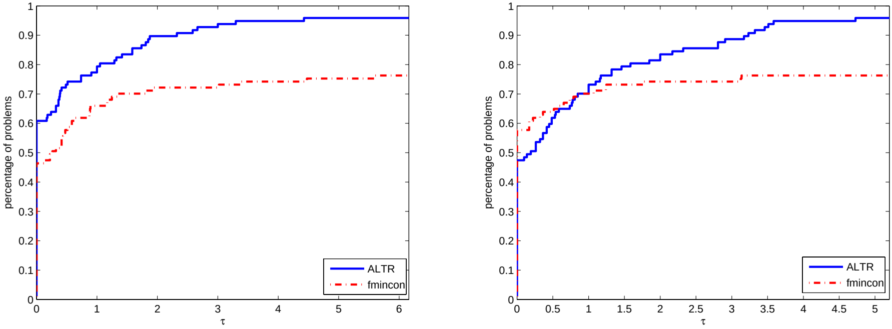


Figure: Comparison between ALTR and fmincon on medium problems: function evaluations (left) and gradient evaluations (right)

Numerical Results Imply

- ALTR performs slightly better than LANCELOT
  - 65%(obtained from $\rho_s(0)$) based on numbers of function evaluations
  - 70% based on numbers of gradient evaluations

- ALTR is more effective and efficient than *fmincon*.

Conclusions

- propose a trust region method motivated by augmented Lagrangian function.
- minimize an approximation of the augmented Lagrangian function within a trust region.
- introduce a new technique for updating the penalty parameters
- a new condition for deciding whether the Lagrange multiplier should be updated.
- Theoretical analysis on convergence properties
- numerical results:
  (comparing with LANCELOT and *fmincon* from Matlab Optimization Toolbox)

Happy Birthday to Philippe Toint!