A Trust Funnel Algorithm for Nonconvex Equality Constrained Optimization with $\mathcal{O}(\epsilon^{-3/2})$ Complexity

Mohammadreza Samadi, Lehigh University

joint work with

Frank E. Curtis (stand-in presenter), Lehigh University

OP17 — Vancouver, British Columbia, Canada

24 May 2017

Outline

Outline

Motivation

Proposed Algorithm

Theoretical Results

Numerical Results

Summary

Introduction

Consider nonconvex equality constrained optimization problems of the form

$$\min_{x \in \mathbb{R}^n} \; f(x)$$
$$\text{s.t. } c(x) = 0,$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $c : \mathbb{R}^n \to \mathbb{R}^m$ are twice continuously differentiable.

- We are interested in algorithm worst-case iteration / evaluation complexity.
- Constraints are not necessarily linear! (No projection-based algorithms.)

Algorithms

Sequential Quadratic Optimization (SQP) / Newton's Method

Trust Funnel; Gould & Toint (2010)

Short-Step ARC; Cartis, Gould, & Toint (2013)

Algorithms

Sequential Quadratic Optimization (SQP) / Newton's Method

- Global convergence: globally convergent (line search or trust region)

Trust Funnel; Gould & Toint (2010)

- Global convergence: globally convergent

Short-Step ARC; Cartis, Gould, & Toint (2013)

- Global convergence: globally convergent

Algorithms

Sequential Quadratic Optimization (SQP) / Newton's Method

- Global convergence: globally convergent (line search or trust region)
- Worst-case complexity: No proved bound

Trust Funnel; Gould & Toint (2010)

- Global convergence: globally convergent
- Worst-case complexity: No proved bound

Short-Step ARC; Cartis, Gould, & Toint (2013)

- Global convergence: globally convergent
- Worst-case complexity: $\mathcal{O}(\epsilon^{-3/2})$ (simplified; more later)

Trust region vs. ARC

Trust Region	ARC
1: Solve to compute s_k:	1: Solve to compute s_k:
$\min_{s\in\mathbb{R}^n} \; q_k(s) := f_k + g_k^T s + \frac{1}{2} s^T H_k s$	$\min_{s\in\mathbb{R}^n} \; c_k(s) := f_k + g_k^T s + \frac{1}{2} s^T H_k s + \frac{1}{3}\sigma_k \\|s\\|_2^3$
s.t. $\\|s\\|_2 \leq \delta_k$ (dual: λ_k)	
2: Compute ratio:	2: Compute ratio:
$\rho_k^q \leftarrow \frac{f_k - f(x_k+s_k)}{f_k - q_k(s_k)}$	$\rho_k^c \leftarrow \frac{f_k - f(x_k+s_k)}{f_k - c_k(s_k)}$
3: Update radius:	3: Update regularization:
$\rho_k^q \geq \eta$: accept and $\delta_k \nearrow$	$\rho_k^c \geq \eta$: accept and $\sigma_k \searrow$
$\rho_k^q < \eta$: reject and $\delta_k \searrow$	$\rho_k^c < \eta$: reject and $\sigma_k \nearrow$

Trust region vs. ARC: Subproblem solution correspondence

Trust Region

1: Solve to compute s_k:

$$\min_{s\in\mathbb{R}^n} \; q_k(s) := f_k + g_k^T s + \tfrac{1}{2} s^T H_k s \quad \text{s.t. } \|s\|_2 \le \delta_k \;\; (\text{dual: } \lambda_k)$$

2: Compute ratio:

$$\rho_k^q \leftarrow \frac{f_k - f(x_k+s_k)}{f_k - q_k(s_k)}$$

3: Update radius:

$\rho_k^q \ge \eta$: accept and $\delta_k \nearrow$

$\rho_k^q < \eta$: reject and $\delta_k \searrow$

$\sigma_k = \frac{\lambda_k}{\delta_k}$ →

← $\delta_k = \|s_k\|_2$

ARC

1: Solve to compute s_k:

$$\min_{s\in\mathbb{R}^n} \; c_k(s) := f_k + g_k^T s + \tfrac{1}{2} s^T H_k s + \tfrac{1}{3}\sigma_k \|s\|_2^3$$

2: Compute ratio:

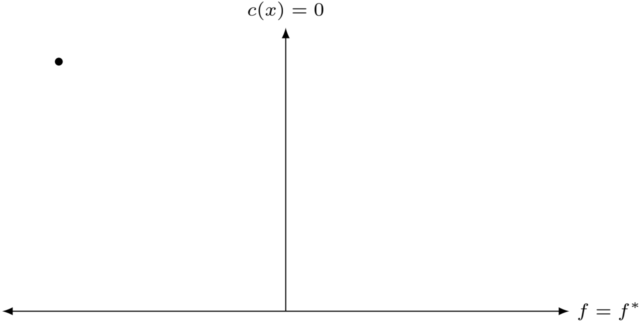
$$\rho_k^c \leftarrow \frac{f_k - f(x_k+s_k)}{f_k - c_k(s_k)}$$

3: Update regularization:

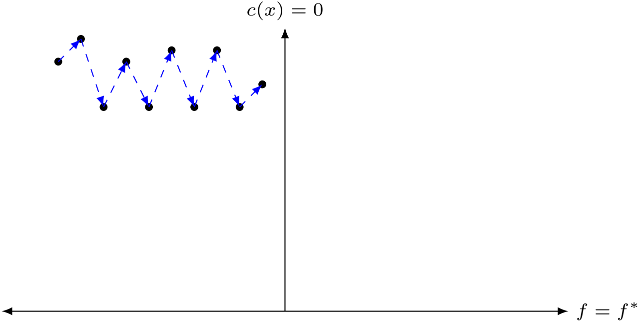
$\rho_k^c \ge \eta$: accept and $\sigma_k \searrow$

$\rho_k^c < \eta$: reject and $\sigma_k \nearrow$

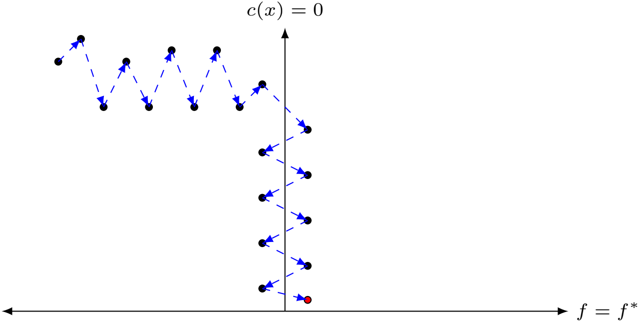
Short-Step ARC



Short-Step ARC



Short-Step ARC



Main concern

- Completely ignores the objective function during the first phase
- Question: Can we do better?

Contributions

- Completely ignores the objective function during the first phase
- Question: Can we do better?
- Yes!(?)
- First, consider `TRACE` method for unconstrained nonconvex optimization
  - FEC, D. P. Robinson, M. Samadi, "A trust region algorithm with a worst-case iteration complexity of $\mathcal{O}(\epsilon^{-3/2})$ for nonconvex optimization," *Mathematical Programming*, 162(1–2), 2017.
- Second, rather than two-phase approach that ignores objective in phase 1, wrap in a **trust funnel** framework that observes objective in both phases.

Contributions

- Completely ignores the objective function during the first phase
- Question: Can we do better?
- Yes!(?)
- First, consider `TRACE` method for unconstrained nonconvex optimization
  - FEC, D. P. Robinson, M. Samadi, "A trust region algorithm with a worst-case iteration complexity of $\mathcal{O}(\epsilon^{-3/2})$ for nonconvex optimization," *Mathematical Programming*, 162(1–2), 2017.
- Second, rather than two-phase approach that ignores objective in phase 1, wrap in a **trust funnel** framework that observes objective in both phases.

Why trust funnel?

- Do not know, in general, how to bound number of updates to a penalty parameter and/or updates of filter entries!
- Trust funnel: "driving factor" is reducing constraint violation.

Outline

SQP "core"

Ideally, given x_k, find s_k as a solution of

$$\min_{s\in\mathbb{R}^n} \; f_k + g_k^T s + \tfrac{1}{2} s^T H_k s$$
$$\text{s.t. } c_k + J_k s = 0$$

Issues:

- H_k (Hessian of Lagrangian) might not be positive definite over $\text{Null}(J_k)$.
- Trust region!. . . but constraints might be incompatible.

Trust funnel basics

Step decomposition approach:

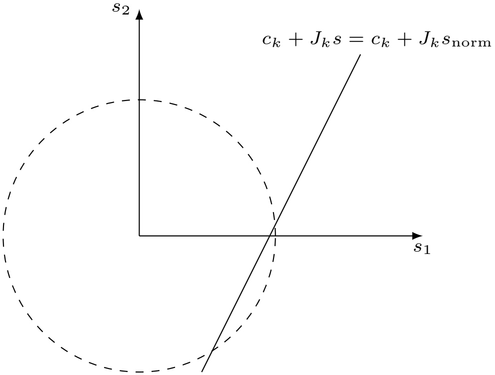
- First, compute a *normal step* toward minimizing constraint violation:

$$v(x) = \tfrac{1}{2}\|c(x)\|_2^2 \quad \Longrightarrow \quad \begin{cases} \min\limits_{s_{\text{norm}} \in \mathbb{R}^n} & m_k^v(s_{\text{norm}}) \\ \quad\text{s.t.} & \|s_{\text{norm}}\|_2 \leq \delta_k^v \end{cases}$$

- Second, compute multipliers λ_k (or take from previous iteration).
- Third, compute a *tangential step* toward optimality:

$$\begin{aligned} \min_{s_{\text{tang}} \in \mathbb{R}^n} \ & m_k^f(s_{\text{norm}} + s_{\text{tang}}) \\ \text{s.t. } & J_k s_{\text{tang}} = 0, \quad \|s_{\text{norm}} + s_{\text{tang}}\|_2 \leq \delta_k^f. \end{aligned}$$

Tangential step



Main idea

Two-phase method combining trust funnel and TRACE.

- Trust funnel for globalization
- TRACE for good complexity bounds

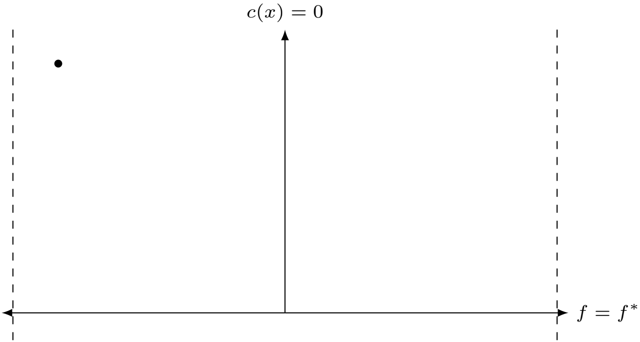
Phase 1 towards feasibility, two types of iterations:

- F-ITERATIONs improve objective and reduce constraint violation.
- V-ITERATIONs reduce constraint violation.

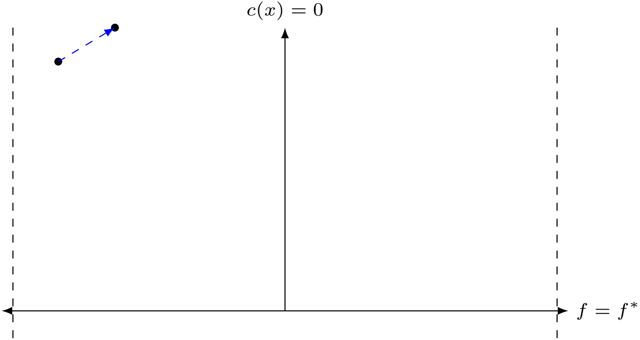
Our algorithm vs. basic trust funnel

- modified F-ITERATION conditions and a different funnel updating procedure
- uses TRACE instead of tradition trust region ideas (for radius updates)
- after getting relatively feasible, switches to "phase 2".

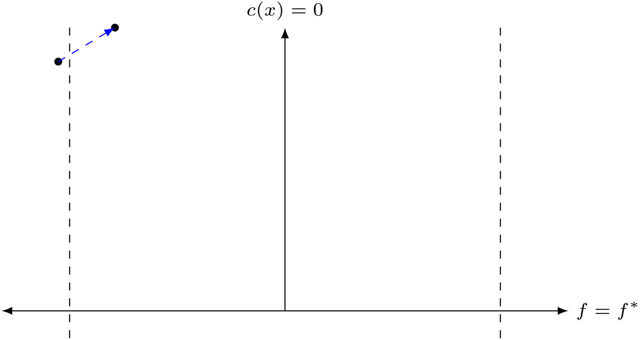
Illustration



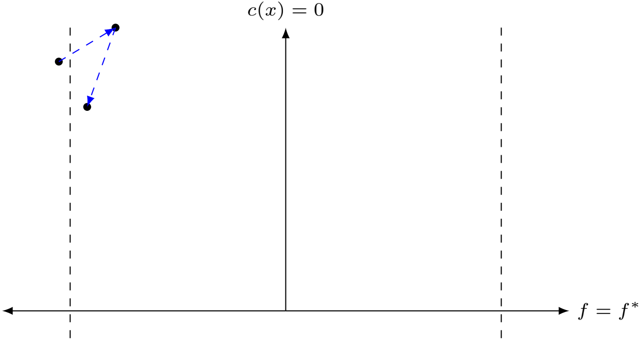
Illustration



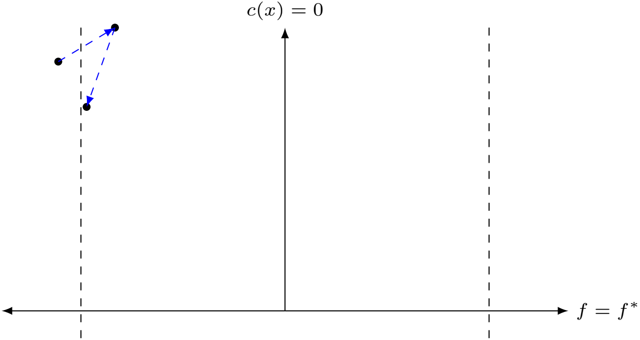
Illustration



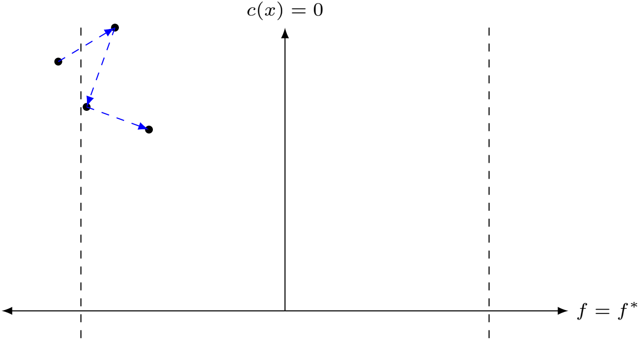
Illustration



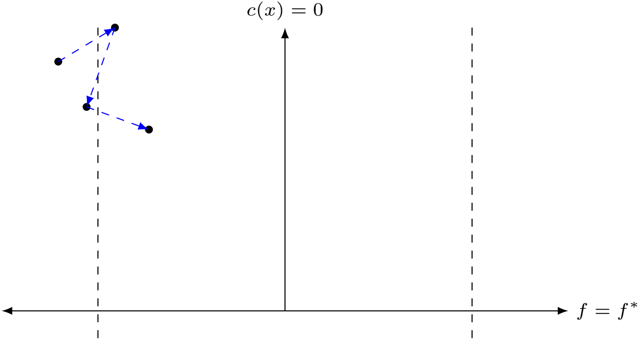
Illustration



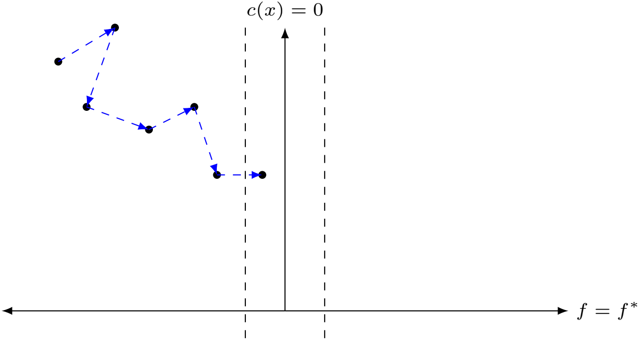
Illustration



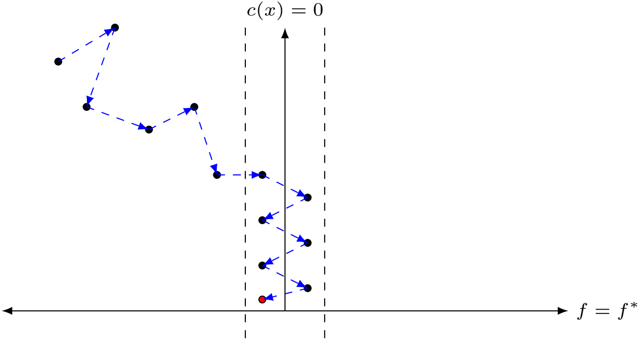
Illustration



Illustration



Illustration



Outline

Phase 1

Recall that $\nabla v(x) = J(x)^T c(x)$ and define the iteration index set

$$\mathcal{I} := \{k \in \mathbb{N} : \|J_k^T c_k\|_2 > \epsilon_v\}.$$

Theorem 1

For any $\epsilon_v \in (0, \infty)$, the cardinality of $\mathcal{I}$ is at most $K(\epsilon_v)$, which accounts for

- $\mathcal{O}(\epsilon_v^{-3/2})$ *successful steps and*
- *finite contraction and expansion steps between successful steps.*

Hence, the cardinality of $\mathcal{I}$ is $K(\epsilon_v) = \mathcal{O}(\epsilon_v^{-3/2})$.

Phase 1

Recall that $\nabla v(x) = J(x)^T c(x)$ and define the iteration index set

$$\mathcal{I} := \{k \in \mathbb{N} : \|J_k^T c_k\|_2 > \epsilon_v\}.$$

Theorem 1

For any $\epsilon_v \in (0, \infty)$, the cardinality of $\mathcal{I}$ is at most $K(\epsilon_v)$, which accounts for

- $\mathcal{O}(\epsilon_v^{-3/2})$ *successful steps and*
- *finite contraction and expansion steps between successful steps.*

Hence, the cardinality of $\mathcal{I}$ is $K(\epsilon_v) = \mathcal{O}(\epsilon_v^{-3/2})$.

Corollary 2

If $\{J_k\}$ have full row rank with singular values bounded below by $\zeta \in (0, \infty)$, then

$$\mathcal{I}_c := \{k \in \mathbb{N} : \|c_k\|_2 > \epsilon_v/\zeta\}$$

has cardinality $\mathcal{O}(\epsilon_v^{-3/2})$.

Phase 2

Options for phase 2:

- trust funnel method (no complexity guarantees) or
- "target-following" approach similar to Short-Step ARC to minimize

$$\Phi(x,t) = \|c(x)\|_2^2 + |f(x) - t|^2$$

Theorem 3

For $\epsilon_f \in (0, \epsilon_v^{1/3}]$, *the number of iterations until*

$$\|g_k + J_k^T y_k\|_2 \le \epsilon_f \|(y_k, 1)\|_2 \quad or \quad \|J_k^T c_k\|_2 \le \epsilon_f \|c_k\|_2$$

is $\mathcal{O}(\epsilon_f^{-3/2}\epsilon_v^{-1/2})$.

Same complexity as Short-Step ARC:

- If $\epsilon_f = \epsilon_v^{2/3}$, then overall $\mathcal{O}(\epsilon_v^{-3/2})$ (though loose KKT tolerance).
- If $\epsilon_f = \epsilon_v$, then overall $\mathcal{O}(\epsilon_v^{-2})$.

Outline

Motivation

Proposed Algorithm

Theoretical Results

Numerical Results

Summary

Implementation

Matlab implementation:

- Phase 1: our algorithm vs. one doing V-ITERATIONs only
- Phase 2: trust funnel method [Curtis, Gould, Robinson, & Toint (2016)]

Termination conditions:

- Phase 1:

$$\|c_k\|_\infty \leq 10^{-6}\max\{\|c_0\|_\infty, 1\} \ \text{ or } \ \begin{cases} \|J_k^T c_k\|_\infty \leq 10^{-6}\max\{\|J_0^T c_0\|_\infty, 1\} \\ \text{and } \|c_k\|_\infty > 10^{-3}\max\{\|c_0\|_\infty, 1\} \end{cases}$$

- Phase 2:

$$\|g_k + J_k^T y_k\|_\infty \leq 10^{-6}\max\{\|g_0 + J_0^T y_0\|_\infty, 1\}.$$

Test set

Equality constrained problems (190) from the CUTEst test set:

78	constant (or null) objective
60	time limit (1 hour)
13	feasible initial point
3	infeasible phase 1
2	function evaluation error
1	small stepsizes (less than 10^{-20})

Remaining set consists of 33 problems.

			TF						TF-V-ONLY				
			Phase 1				Phase 2		Phase 1			Phase 2	
Problem	n	m	#V	#F	f	$\|g + J^T y\|$	#V	#F	#V	f	$\|g + J^T y\|$	#V	#F
BT1	2	1	4	0	-8.02e-01	+4.79e-01	0	139	4	-8.00e-01	+7.04e-01	7	136
BT10	2	2	10	0	-1.00e+00	+5.39e-04	1	0	10	-1.00e+00	+6.74e-05	1	0
BT11	5	3	6	1	+8.25e-01	+4.84e-03	2	0	1	+4.55e+04	+2.57e+04	16	36
BT12	5	3	12	1	+6.19e+00	+1.18e-05	0	0	16	+3.34e+01	+4.15e+00	4	8
BT2	3	1	22	8	+1.45e+03	+3.30e+02	3	12	21	+6.14e+04	+1.82e+04	0	40
BT3	5	3	1	0	+4.09e+00	+6.43e+02	1	0	1	+1.01e+05	+8.89e+02	0	1
BT4	3	2	1	0	-1.86e+01	+1.00e+01	20	12	1	-1.86e+01	+1.00e+01	20	12
BT5	3	2	15	2	+9.62e+02	+2.80e+00	14	2	8	+9.62e+02	+3.83e-01	3	1
BT6	5	2	11	45	+2.77e-01	+4.64e-02	1	0	14	+5.81e+02	+4.50e+02	5	59
BT7	5	3	15	6	+1.31e+01	+5.57e+00	5	1	12	+1.81e+01	+1.02e+01	19	28
BT8	5	2	50	26	+1.00e+00	+7.64e-04	1	1	10	+2.00e+00	+2.00e+00	1	97
BT9	4	2	11	1	-1.00e+00	+8.56e-05	1	0	10	-9.69e-01	+2.26e-01	5	1
BYRDSPHR	3	2	29	2	-4.68e+00	+1.28e-05	0	0	19	-5.00e-01	+1.00e+00	16	5
CHAIN	800	401	9	0	+5.12e+00	+2.35e-04	3	20	9	+5.12e+00	+2.35e-04	3	20
FLT	2	2	15	4	+2.68e+10	+3.28e+05	0	13	19	+2.68e+10	+3.28e+05	0	17
GENHS28	10	8	1	0	+9.27e-01	+5.88e+01	0	0	1	+2.46e+03	+9.95e+01	0	1
HS100LNP	7	2	16	2	+6.89e+02	+1.74e+01	4	1	5	+7.08e+02	+1.93e+01	14	3
HS111LNP	10	3	9	1	-4.78e+01	+4.91e-06	2	0	10	-4.62e+01	+7.49e-01	10	1
HS27	3	1	2	0	+8.77e+01	+2.03e+02	3	5	1	+2.54e+01	+1.41e+02	11	34
HS39	4	2	11	1	-1.00e+00	+8.56e-05	1	0	10	-9.69e-01	+2.26e-01	5	1
HS40	4	3	4	0	-2.50e-01	+1.95e-06	0	0	3	-2.49e-01	+3.35e-02	2	1
HS42	4	2	4	1	+1.39e+01	+3.94e-04	1	0	1	+1.50e+01	+2.00e+00	3	1
HS52	5	3	1	0	+5.33e+00	+1.54e+02	1	0	1	+8.07e+03	+4.09e+02	0	1
HS6	2	1	1	0	+4.84e+00	+1.56e+00	32	136	1	+4.84e+00	+1.56e+00	32	136
HS7	2	1	7	1	-2.35e-01	+1.18e+00	7	2	8	+3.79e-01	+1.07e+00	5	2
HS77	5	2	13	30	+2.42e-01	+1.26e-02	0	0	17	+5.52e+02	+4.54e+02	3	11
HS78	5	3	6	0	-2.92e+00	+3.65e-04	1	0	10	-1.79e+00	+1.77e+00	2	30
HS79	5	3	13	21	+7.88e-02	+5.51e-02	0	2	10	+9.70e+01	+1.21e+02	0	24
MARATOS	2	1	4	0	-1.00e+00	+8.59e-05	1	0	3	-9.96e-01	+9.02e-02	2	1
MSS3	2070	1981	12	0	-4.99e+01	+2.51e-01	50	0	12	-4.99e+01	+2.51e-01	50	0
MWRIGHT	5	3	17	6	+2.31e+01	+5.78e-05	1	0	7	+5.07e+01	+1.04e+01	12	20
ORTHREGB	27	6	10	15	+7.02e-05	+4.23e-04	0	6	10	+2.73e+00	+1.60e+00	0	10
SPIN2OP	102	100	57	18	+2.04e-08	+2.74e-04	0	1	time	+1.67e+01	+3.03e-01	time	time

Problem	n	m	TF						TF-V-ONLY				
			Phase 1				Phase 2		Phase 1			Phase 2	
			#V	#F	f	$\|g + J^T y\|$	#V	#F	#V	f	$\|g + J^T y\|$	#V	#F
BT11	5	3	6	1	+8.25e-01	+4.84e-03	2	0	1	+4.55e+04	+2.57e+04	16	36
BT7	5	3	15	6	+1.31e+01	+5.57e+00	5	1	12	+1.81e+01	+1.02e+01	19	28

Summary of results

Our algorithm, at the end of phase 1

- for 26 problems, reaches a smaller function value
- for 6 problems, reaches the same function value

Total number of iterations of our algorithm

- for 18 problems is smaller
- for 8 problems is equal

Outline

Summary

- Proposed an algorithm for equality constrained optimization
- Trust funnel algorithm with improved complexity properties
- Promising performance in practice based on our preliminary experiment
- A step toward practical algorithms with good iteration complexity

⋆ F. E. Curtis, D. P. Robinson, and M. Samadi.
Complexity Analysis of a Trust Funnel Algorithm for Equality Constrained Optimization.
Technical Report 16T-013, COR@L Laboratory, Department of ISE, Lehigh University, 2016.