A Matrix-free Method for Equality Constrained Optimization Problems with Rank Deficient Jacobians

Frank E. Curtis
New York University

involving joint work with
Richard H. Byrd, Jorge Nocedal, and Andreas Wächter

Copper Mountain, 2008

Outline

Outline

Equality constrained optimization

We consider *very large* 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 $c : \mathbb{R}^n \to \mathbb{R}^t$ are smooth functions

- First, we describe a matrix-free primal-dual method for nice cases
- Then, we show how we handle (near) rank deficiency
- Assume strict convexity here, but we can handle non-convexity as well

First-order optimality

Defining the Lagrangian

$$\mathcal{L}(x, \lambda) \triangleq f(x) + \lambda^T c(x)$$

we are interested in finding a first-order optimal point; i.e., one satisfying

$$\nabla \mathcal{L} = \begin{bmatrix} g(x) + A(x)^T \lambda \\ c(x) \end{bmatrix} = 0$$

where $g(x)$ is the gradient of $f(x)$ and $A(x)$ is the Jacobian of $c(x)$

Note: if the problem is infeasible, we would like to at least guarantee convergence toward a stationary point of the feasibility measure

$$\varphi(x) = \|c(x)\|;$$

that is, one satisfying

$$A(x)^T c(x) = 0$$

Method of choice: Newton/SQP

A Newton iteration from the point (x_k, λ_k) has the form

$$\begin{bmatrix} W(x_k, \lambda_k) & A(x_k)^T \\ A(x_k) & 0 \end{bmatrix} \begin{bmatrix} d_k \\ \delta_k \end{bmatrix} = - \begin{bmatrix} g(x_k) + A(x_k)^T \lambda_k \\ c(x_k) \end{bmatrix}$$

where $W(x_k, \lambda_k) \approx \nabla^2_{xx} \mathcal{L}(x_k, \lambda_k)$, which is equivalent to solving the Sequential Quadratic Programming (SQP) subproblem

$$\begin{aligned} \min_{d \in \mathbb{R}^n} \; & f(x_k) + g(x_k)^T d + \tfrac{1}{2} d^T W(x_k, \lambda_k) d \\ \text{s.t. } \; & c(x_k) + A(x_k) d = 0 \end{aligned}$$

Note: step may be arbitrarily large in norm if A is ill-conditioned, and step computation may not even be defined if $rank(A) < t$

Globalization with an exact penalty function

Algorithm outline: for $k = 0, 1, 2, \ldots$

- ... evaluate f_k, g_k, c_k, A_k, and W_k
- ... solve the *primal-dual* equations

$$\begin{bmatrix} W_k & A_k^T \\ A_k & 0 \end{bmatrix} \begin{bmatrix} d_k \\ \delta_k \end{bmatrix} = -\begin{bmatrix} g_k + A_k^T \lambda_k \\ c_k \end{bmatrix}$$

$$\min_{d \in \mathbb{R}^n} \; f(x_k) + g(x_k)^T d + \tfrac{1}{2} d^T W(x_k, \lambda_k) d \quad \text{s.t. } c(x_k) + A(x_k) d = 0$$

- ... set the penalty parameter π_k
- ... perform a line search for the merit function

$$\phi(x; \pi_k) \triangleq f(x) + \pi_k \|c(x)\|$$

to find $\alpha_k \in (0, 1]$ satisfying the Armijo condition

$$\phi(x_k + \alpha_k d_k; \pi_k) \leq \phi(x_k; \pi_k) + \eta \alpha_k D\phi(d_k; \pi_k)$$

Outline

Working with matrices may be impractical

$$\begin{bmatrix} W_k & A_k^T \\ A_k & 0 \end{bmatrix} \begin{bmatrix} d_k \\ \delta_k \end{bmatrix} = - \begin{bmatrix} g_k + A_k^T \lambda_k \\ c_k \end{bmatrix}$$

What if...

- A_k, A_k^T, and W_k cannot be computed explicitly?
- A_k, A_k^T, and W_k cannot be stored?
- the *primal-dual matrix* cannot be factored?
- an iterative method may be more efficient?

If the products $A_k p$, $A_k^T q$, and $W_k y$ can be computed, we have answers...

Iterative step computations

From now on, let us assume that we have an iterative procedure for solving the primal-dual equations, which during each *inner iteration* yields (d_k, δ_k) solving

$$\begin{bmatrix} W_k & A_k^T \\ A_k & 0 \end{bmatrix} \begin{bmatrix} d_k \\ \delta_k \end{bmatrix} = - \begin{bmatrix} g_k + A_k^T \lambda_k \\ c_k \end{bmatrix} + \begin{bmatrix} \rho_k \\ r_k \end{bmatrix}$$

for the residuals (ρ_k, r_k)

- How can we be sure that a given inexact step is *acceptable*?
- How small do the residuals need to be?

Outline

A naïve approach

Algorithm outline: given $0 < \kappa < 1$, for $k = 0, 1, 2, \ldots$

- ... evaluate f_k, g_k, c_k, $A_k^T \lambda_k$
- ... solve the *primal-dual* equations

$$\boxed{\begin{bmatrix} W_k & A_k^T \\ A_k & 0 \end{bmatrix} \begin{bmatrix} d_k \\ \delta_k \end{bmatrix} = - \begin{bmatrix} g_k + A_k^T \lambda_k \\ c_k \end{bmatrix} + \begin{bmatrix} \rho_k \\ r_k \end{bmatrix}}$$

  until $\|(\rho_k, r_k)\| \le \kappa \|(g_k + A_k^T \lambda_k, c_k)\|$

- ... set the penalty parameter π_k
- ... perform a line search to find $\alpha_k \in (0, 1]$ satisfying

$$\phi(x_k + \alpha_k d_k; \pi_k) \le \phi(x_k; \pi_k) + \eta \alpha_k D\phi(d_k; \pi_k)$$

A naïve approach

Algorithm outline: given $0 < \kappa < 1$, for $k = 0, 1, 2, \ldots$

- ... evaluate f_k, g_k, c_k, $A_k^T \lambda_k$
- ... solve the *primal-dual* equations

$$\boxed{\begin{bmatrix} W_k & A_k^T \\ A_k & 0 \end{bmatrix} \begin{bmatrix} d_k \\ \delta_k \end{bmatrix} = - \begin{bmatrix} g_k + A_k^T \lambda_k \\ c_k \end{bmatrix} + \begin{bmatrix} \rho_k \\ r_k \end{bmatrix}}$$

until $\|(\rho_k, r_k)\| \leq \kappa \|(g_k + A_k^T \lambda_k, c_k)\|$

- ... set the penalty parameter π_k
- ... perform a line search to find $\alpha_k \in (0, 1]$ satisfying

$$\phi(x_k + \alpha_k d_k; \pi_k) \leq \phi(x_k; \pi_k) + \eta \alpha_k \underbrace{D\phi(d_k; \pi_k)}_{>0\ \forall \pi ?}$$

κ	2^{-1}	2^{-5}	2^{-10}
% Solved	45%	80%	86%

Optimization, not nonlinear equations

$$\begin{bmatrix} W_k & A_k^T \\ A_k & 0 \end{bmatrix} \begin{bmatrix} d_k \\ \delta_k \end{bmatrix} = - \begin{bmatrix} g_k + A_k^T \lambda_k \\ c_k \end{bmatrix} + \begin{bmatrix} \rho_k \\ r_k \end{bmatrix}$$

$$\min_{d \in \mathbb{R}^n} \ f_k + g_k^T d + \tfrac{1}{2} d^T W_k d \quad \text{s.t.} \ c_k + A_k d = 0$$

Take (d_k, δ_k) and...

- ... "forget" about it being an inexact Newton step
- ... "forget" about it being an approximate SQP solution

We want a technique for determining if (d_k, δ_k) is acceptable that...

- ... allows for possibly very inexact solutions to Newton's equations
- ... integrates both step computation and step selection to solve the optimization problem

Central idea: Sufficient Model Reductions

Modern optimization algorithms work with models.

Take the penalty function

$$\phi(x;\pi) \triangleq f(x) + \pi\|c(x)\|$$

and consider the model

$$m_k(d;\pi) \triangleq f_k + g_k^T d + \pi\|c_k + A_k d\|$$

The reduction in m_k attained by d_k is computed easily as

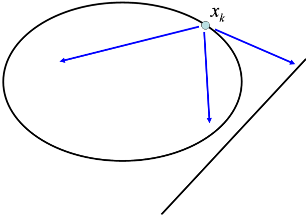
$$\begin{aligned}\Delta m_k(d_k;\pi) &\triangleq m_k(0;\pi) - m_k(d_k;\pi)\\ &= -g_k^T d_k + \pi(\|c_k\| - \|r_k\|)\end{aligned}$$

and yields

$$D\phi(d_k;\pi) \leq -\Delta m_k(d_k;\pi)$$

Main tool: "SMART" Tests

We develop two types of
Sufficient Merit function Approximation Reduction Termination Tests.

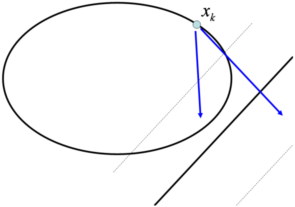


Termination Test I: A sufficient model reduction is attained for π_{k-1} (i.e., the most recent penalty parameter value):

$$\Delta m_k(d_k;\pi_{k-1}) = -g_k^T d_k + \pi_{k-1}(\|c_k\| - \|r_k\|) \gg 0$$

Main tool: "SMART" Tests

We develop two types of
Sufficient Merit function Approximation Reduction Termination Tests.



Termination Test II: A sufficient reduction in the constraint model is attained for some $\epsilon \in (0, 1)$

$$\|r_k\| \leq \epsilon \|c_k\|$$

Step acceptance criteria:

Model Reduction Condition. A step (d_k, δ_k) is acceptable if and only if

$$\Delta m_k(d_k; \pi_k) \geq \frac{1}{2} d_k^T W_k d_k + \sigma \pi_k \max\{\|c_k\|, \|c_k + A_k d_k\| - \|c_k\|\}$$

for some $\sigma \in (0,1)$ and an appropriate $\pi_k > 0$.

Termination Test I. For some $\sigma \in (0,1)$ and $\pi_k = \pi_{k-1}$ the Model Reduction Condition is satisfied and for some $\kappa \in (0,1)$ we have

$$\left\| \begin{bmatrix} \rho_k \\ r_k \end{bmatrix} \right\| \leq \kappa \left\| \begin{bmatrix} g_k + A_k^T \lambda_k \\ c_k \end{bmatrix} \right\|$$

Termination Test II. For some $\epsilon \in (0,1)$ and $\beta > 0$ we have

$$\|r_k\| \leq \epsilon \|c_k\| \quad \text{and} \quad \|\rho_k\| \leq \beta \|c_k\|$$

and we set

$$\pi_k \geq \frac{g_k^T d_k + \frac{1}{2} d_k^T W_k d_k}{(1-\tau)(\|c_k\| - \|r_k\|)} \qquad \text{for } \tau \in (0,1)$$

Inexact SQP with SMART Tests[1]

Algorithm outline: for $k = 0, 1, 2 \dots$

- ... evaluate f_k, g_k, c_k, $A_k^T \lambda_k$
- ... solve the *primal-dual* equations

$$\boxed{\begin{bmatrix} W_k & A_k^T \\ A_k & 0 \end{bmatrix} \begin{bmatrix} d_k \\ \delta_k \end{bmatrix} = - \begin{bmatrix} g_k + A_k^T \lambda_k \\ c_k \end{bmatrix} + \begin{bmatrix} \rho_k \\ r_k \end{bmatrix}}$$

until Termination Test I or II holds

- ... set the penalty parameter π_k
- ... perform a line search to find $\alpha_k \in (0, 1]$ satisfying

$$\phi(x_k + \alpha_k d_k; \pi_k) \leq \phi(x_k; \pi_k) - \eta \alpha_k \Delta m_k(d_k; \pi_k)$$

[1] R. H. Byrd, F. E. Curtis, and J. Nocedal, "An Inexact SQP Method for Equality Constrained Optimization," to appear in SIAM Journal on Optimization.

Outline

(Near) Rank-deficient Jacobians

If at any point the Jacobian A of c is ill-conditioned or rank deficient, the Newton system

$$\begin{bmatrix} W(x_k,\lambda_k) & A(x_k)^T \\ A(x_k) & 0 \end{bmatrix} \begin{bmatrix} d_k \\ \delta_k \end{bmatrix} = - \begin{bmatrix} g(x_k) + A(x_k)^T\lambda_k \\ c(x_k) \end{bmatrix}$$

and the SQP subproblem

$$\begin{aligned} \min_{d\in\mathbb{R}^n} \ & f(x_k) + g(x_k)^T d + \tfrac{1}{2} d^T W(x_k,\lambda_k) d \\ \text{s.t. } \ & c(x_k) + A(x_k) d = 0 \end{aligned}$$

may not be well-defined or may lead to very long steps (i.e., $\|d_k\| \gg 0$, $\alpha_k \approx 0$, and algorithm may stall)

Regularizing the constraint model with trust regions

We decompose the step by first considering the trust region subproblem

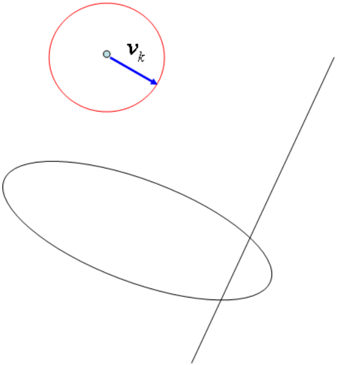
$$\min_{v \in \mathbb{R}^n} \ \tfrac{1}{2}\|c_k + A_k v\|^2$$
$$\text{s.t. } \|v\| \leq \Omega_k$$

Notice that this subproblem fits well within our context of matrix-free optimization; e.g., apply CG/LSQR with Steihaug-Toint stop tests

Trust regions

The trust region keeps us in a local region of the search space:

$$\min_{v \in \mathbb{R}^n} \frac{1}{2}\|c_k + A_k v\|^2$$
$$\text{s.t. } \|v\| \leq \Omega_k$$



Trust regions

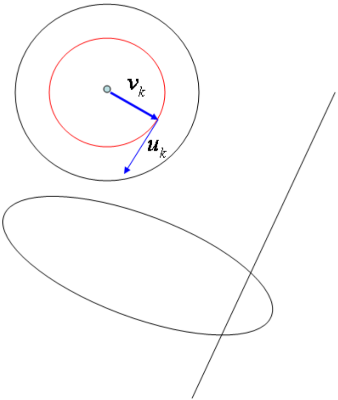
Once v is computed, we could consider computing a step toward optimality within a larger trust region:

$$\min_{u \in \mathbb{R}^n} \; (g_k + W_k v_k)^T u + \tfrac{1}{2} u^T W_k u$$
$$\text{s.t. } A_k u = 0, \quad \|u\| \le \Omega'_k,$$

but then we may need

$$Z_k \quad \text{s.t.} \quad A_k Z_k \approx 0$$

or to (approximately) project vectors onto the null space of A_k



Trust regions only for v!

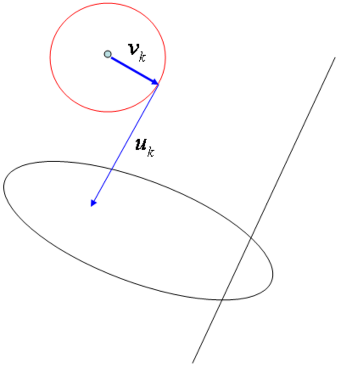
Instead, we set no trust region for u:

$$\min_{u \in \mathbb{R}^n} \; (g_k + W_k v_k)^T u + \frac{1}{2} u^T W_k u$$
$$\text{s.t. } A_k u = 0$$

which, with $d_k = v_k + u_k$, has the same solutions as

$$\begin{bmatrix} W_k & A_k^T \\ A_k & 0 \end{bmatrix} \begin{bmatrix} d_k \\ \delta_k \end{bmatrix} = \begin{bmatrix} -(g_k + A_k^T \lambda_k) \\ A_k v_k \end{bmatrix}$$

Notice that this system is consistent (though perhaps (near) singular)



Setting the trust region radius

In fact, we propose a very specific form for the trust region radius:

$$\min_{v\in\mathbb{R}^n} \tfrac{1}{2}\|c_k + A_k v\|^2$$
$$\text{s.t. } \|v\| \leq \omega\|A_k^T c_k\|$$

for a given *constant* $\omega > 0$

- We incorporate problem information in the right-hand-side (recall that a stationary point for the feasibility measure has $A^T c = 0$)
- The radius is set dynamically without a heuristic update
- ω should be set to correspond to the reciprocal of the smallest allowable singular value of A_k

Step acceptance criteria:[2]

Tangential Component Condition. The component u_k must satisfy

$$\|u_k\| \leq \psi\|v_k\| \quad \text{or} \quad (g_k + W_k v_k)^T u_k + \tfrac{1}{2} u_k^T W_k u_k \leq 0$$

Model Reduction Condition. A step (d_k, δ_k) is acceptable if and only if

$$\Delta m_k(d_k; \pi_k) \geq \tfrac{1}{2} u_k^T W_k u_k + \sigma\pi_k(\|c_k\| - \|c_k + A_k v_k\|)$$

for some $\sigma \in (0,1)$ and an appropriate $\pi_k > 0$.

Termination Test I. For some $\sigma \in (0,1)$ and $\pi_k = \pi_{k-1}$ the Tangential Component Condition holds, the Model Reduction Condition is satisfied, and for some $\kappa \in (0,1)$ we have

$$\left\| \begin{bmatrix} \rho_k \\ r_k \end{bmatrix} \right\| \leq \kappa \min \left\{ \left\| \begin{bmatrix} g_k + A_k^T \lambda_k \\ A_k v_k \end{bmatrix} \right\|, \left\| \begin{bmatrix} g_{k-1} + A_{k-1}^T \lambda_k \\ A_{k-1} v_{k-1} \end{bmatrix} \right\| \right\}$$

Termination Test II. For some $\epsilon \in (0,1)$ and $\beta > 0$, the Tangential Component Condition holds and we have

$$\begin{aligned} \|c_k\| - \|c_k + A_k d_k\| &\geq \epsilon(\|c_k\| - \|c_k + A_k v_k\|) \\ \text{and} \quad \|\rho_k\| &\leq \beta(\|c_k\| - \|c_k + A_k v_k\|), \\ \text{and we set} \quad \pi_k &\geq (g_k^T d_k + \tfrac{1}{2} u_k^T W_k u_k)/((1-\tau)(\|c_k\| - \|c_k + A_k d_k\|)) \end{aligned}$$

[2] F. E. Curtis, J. Nocedal, and A. Wächter, in preparation.

Outline

Main result

<u>Assumptions</u>: The generated sequence $\{x_k, \lambda_k\}$ is contained in a convex set over which f and c and their first derivatives are bounded, and the iterative linear system solver can solve the primal-dual equations to an arbitrary accuracy

<u>Theorem</u>: If all limit points satisfy the linear independence constraint qualification (LICQ), then $\{\pi_k\}$ is bounded and

$$\lim_{k \to \infty} \left\| \begin{bmatrix} g_k + A_k^T \lambda_{k+1} \\ c_k \end{bmatrix} \right\| = 0$$

Otherwise,

$$\lim_{k \to \infty} \left\| A_k^T c_k \right\| = 0$$

and if $\{\pi_k\}$ is bounded then

$$\lim_{k \to \infty} \left\| g_k + A_k^T \lambda_{k+1} \right\| = 0$$

Brief overview of analysis

- The step length (d_k, v_k, u_k) is explicitly or implicitly controlled...
- The reduction in the model of the penalty function satisfies

$$\Delta m_k(d_k; \pi_k) \geq \gamma(\|u_k\|^2 + \pi_k \|A_k^T c_k\|^2)$$

- In particular

$$\Delta m_k(d_k; \pi_k) \geq \gamma' \|A_k^T c_k\|^2 \Rightarrow \lim_{k \to \infty} \|A_k^T c_k\| = 0$$

- If $\{\pi_k\}$ remains bounded (guaranteed if LICQ holds), then

$$\lim_{k \to \infty} \left\| g_k + A_k^T \lambda_{k+1} \right\| = 0,$$

and otherwise $\pi \to \infty$

Outline

Implementation details

We use MINRES to solve the primal-dual equations

$$\begin{bmatrix} W_k & A_k^T \\ A_k & 0 \end{bmatrix} \begin{bmatrix} d_k \\ \delta_k \end{bmatrix} = \begin{cases} -\begin{bmatrix} g_k + A_k^T \lambda_k \\ c_k \end{bmatrix} \\ -\begin{bmatrix} g_k + A_k^T \lambda_k \\ -A_k v_k \end{bmatrix} \end{cases}$$

and LSQR (algebraically equivalent to CG, but with better numerical properties) with Steihaug-Toint stop tests to solve the trust region subproblem

$$\min_{v \in \mathbb{R}^n} \tfrac{1}{2} \|c_k + A_k v\|^2$$
$$\text{s.t. } \|v\| \leq \omega \|A_k^T c_k\|$$

All experiments performed in Matlab

Briefly, the nice case

κ	2^{-1}	2^{-5}	2^{-10}	iSQP
% Solved	45%	80%	86%	100%

Problems with rank-deficiency

Total of 73 problems from the CUTEr collection

- Original and perturbed models have

$$c_1(x) = 0 \quad \text{and} \quad \begin{cases} c_1(x) = 0 \\ c_1(x) - c_1^2(x) = 0 \end{cases}$$

  respectively

- Success rates:

	iSQP	TRINS
Original	95%	100%
Perturbed	46%	93%

- A few of the failures of TRINS was due to the Maratos effect, so second-order correction steps may be beneficial

Conclusion

We have...

- ... focused on a particular class of problems to which contemporary optimization techniques cannot be applied
- ... considered the fundamental question of how to ensure global convergence via a type of inexact SQP/Newton approach
- ... developed a novel methodology where inexact solutions are appraised based on the reductions obtained in linear models of an exact penalty function
- ... extended the algorithm and analysis for cases involving rank deficiency (and nonconvexity)