Negative Curvature and Equality Constrained Optimization

Richard H. Byrd[1] Frank E. Curtis[2] Jorge Nocedal[2]

[1]University of Colorado at Boulder

[2]Northwestern University

Computational Science and Engineering, 2003

Outline

Outline

Problem Formulation

Nonlinear program with equality constraints

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

Sequential Quadratic Programming (SQP)

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

Step can be obtained via primal-dual system

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

Problem Formulation

Nonlinear program with equality constraints

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

Sequential Quadratic Programming (SQP)

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

Step can be obtained via primal-dual system

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

Problem Formulation

Nonlinear program with equality constraints

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

Sequential Quadratic Programming (SQP)

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

Step can be obtained via primal-dual system

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

Algorithm Framework

Line Search SQP Method

- Apply an iterative solver to the primal-dual system

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

  until a stopping condition is satisfied[1]

- Set the penalty parameter π to ensure descent on the penalty function

$$\phi_\pi(x) = f(x) + \pi \|c(x)\|$$

- Perform a backtracking line search to find α_k satisfying the Armijo condition

$$\phi_{\pi_k}(x_k + \alpha_k d_k) \leq \phi_{\pi_k}(x_k) + \eta \alpha_k D\phi_{\pi_k}(d_k)$$

[1] (Byrd, Curtis, Nocedal, 2007)

Algorithm Framework

Line Search SQP Method

- Apply an iterative solver to the primal-dual system

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

until a stopping condition is satisfied[1]

- Set the penalty parameter π to ensure descent on the penalty function

$$\phi_\pi(x) = f(x) + \pi \| c(x) \|$$

- Perform a backtracking line search to find α_k satisfying the Armijo condition

$$\phi_{\pi_k}(x_k + \alpha_k d_k) \leq \phi_{\pi_k}(x_k) + \eta \alpha_k D\phi_{\pi_k}(d_k)$$

[1] (Byrd, Curtis, Nocedal, 2007)

Algorithm Framework

Line Search SQP Method

- Apply an iterative solver to the primal-dual system

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

  until a stopping condition is satisfied[1]

- Set the penalty parameter π to ensure descent on the penalty function

$$\phi_\pi(x) = f(x) + \pi \|c(x)\|$$

- Perform a backtracking line search to find α_k satisfying the Armijo condition

$$\phi_{\pi_k}(x_k + \alpha_k d_k) \leq \phi_{\pi_k}(x_k) + \eta \alpha_k D\phi_{\pi_k}(d_k)$$

[1](Byrd, Curtis, Nocedal, 2007)

Stopping Conditions

Inexactness determined by reductions in ϕ_π

$$\text{mred}_\pi(d) = -g_k^T d - \frac{\omega(d)}{2} d^T W_k d + \pi(\|c_k\| - \|r_k\|)$$

An accepted step (d_k, δ_k) must satisfy

$$\text{mred}_{\pi_k}(d_k) \geq \sigma \pi_k \max\{\|c_k\|, \|r_k\|\} \tag{1}$$

for a given constant $0 < \sigma < 1$ and appropriate π_k

- Stopping Condition I: (1) satisfied for $\pi_k = \pi_{k-1}$
- Stopping Condition II: $\|r_k\| \leq \epsilon \|c_k\|$ for $0 < \epsilon < 1$ and

$$\pi_k \geq \frac{g_k^T d_k + \frac{\omega_k}{2} d_k^T W_k d_k}{(1-\sigma)(\|c_k\| - \|r_k\|)}$$

Stopping Conditions

Inexactness determined by reductions in ϕ_π

$$\text{mred}_\pi(d) = -g_k^T d - \frac{\omega(d)}{2} d^T W_k d + \pi(\|c_k\| - \|r_k\|)$$

An accepted step (d_k, δ_k) must satisfy

$$\text{mred}_{\pi_k}(d_k) \geq \sigma \pi_k \max\{\|c_k\|, \|r_k\|\} \quad (1)$$

for a given constant $0 < \sigma < 1$ and appropriate π_k

- Stopping Condition I: (1) satisfied for $\pi_k = \pi_{k-1}$
- Stopping Condition II: $\|r_k\| \leq \epsilon \|c_k\|$ for $0 < \epsilon < 1$ and

$$\pi_k \geq \frac{g_k^T d_k + \frac{\omega_k}{2} d_k^T W_k d_k}{(1-\sigma)(\|c_k\| - \|r_k\|)}$$

Stopping Conditions

Inexactness determined by reductions in ϕ_π

$$\text{mred}_\pi(d) = -g_k^T d - \frac{\omega(d)}{2} d^T W_k d + \pi(\|c_k\| - \|r_k\|)$$

An accepted step (d_k, δ_k) must satisfy

$$\text{mred}_{\pi_k}(d_k) \geq \sigma \pi_k \max\{\|c_k\|, \|r_k\|\} \tag{1}$$

for a given constant $0 < \sigma < 1$ and appropriate π_k

- Stopping Condition I: (1) satisfied for $\pi_k = \pi_{k-1}$
- Stopping Condition II: $\|r_k\| \leq \epsilon \|c_k\|$ for $0 < \epsilon < 1$ and

$$\pi_k \geq \frac{g_k^T d_k + \frac{\omega_k}{2} d_k^T W_k d_k}{(1-\sigma)(\|c_k\| - \|r_k\|)}$$

Outline

Assumptions for Global Convergence

Global convergence is guaranteed if

$$Z_k^T W_k Z_k \succ 0$$

where Z_k is a basis for the null space of A_k (i.e., $A_k Z_k = 0$), which is known to hold if the primal-dual matrix

$$\begin{bmatrix} W_k & A_k^T \\ A_k & 0 \end{bmatrix}$$

has n positive and m negative eigenvalues

Assumptions for Global Convergence

Global convergence is guaranteed if

$$Z_k^T W_k Z_k \succ 0$$

where Z_k is a basis for the null space of A_k (i.e., $A_k Z_k = 0$), which is known to hold if the primal-dual matrix

$$\begin{bmatrix} W_k & A_k^T \\ A_k & 0 \end{bmatrix}$$

has n positive and m negative eigenvalues

Step Decomposition

Step can be decomposed into a tangential step u_k, lying in the null space of A_k, and a normal step v_k, lying in the range space of A_k^T

$$d_k = u_k + v_k, \quad \text{where} \quad \|d_k\|^2 = \|u_k\|^2 + \|v_k\|^2$$

We claim that if Stopping Condition I or II is satisfied

- ... and
$$\theta \|u_k\|^2 \leq \|v_k\|^2$$
then step is acceptable
- ... and
$$\theta \|u_k\|^2 \leq d_k^T W_k d_k$$
then step is acceptable

Step Decomposition

Step can be decomposed into a tangential step u_k, lying in the null space of A_k, and a normal step v_k, lying in the range space of A_k^T

$$d_k = u_k + v_k, \quad \text{where} \quad \|d_k\|^2 = \|u_k\|^2 + \|v_k\|^2$$

We claim that if Stopping Condition I or II is satisfied

- ... and
$$\theta \|u_k\|^2 \leq \|v_k\|^2$$
then step is acceptable
- ... and
$$\theta \|u_k\|^2 \leq d_k^T W_k d_k$$
then step is acceptable

Observing a Bounded Tangential Step

Observe

$$\begin{aligned} \|v_k\| &\geq \|A_k v_k\|/\|A_k\| \\ &= \|A_k d_k\|/\|A_k\| \end{aligned}$$

and so

$$\theta\|d_k\|^2 \leq \|A_k d_k\|^2/\|A_k\|^2$$

implies

$$\theta\|u_k\|^2 \leq \|v_k\|^2$$

Observing Negative Curvature

Observe

$$\begin{aligned} \|u_k\|^2 = & \ \|d_k\|^2 - \|v_k\|^2 \\ \leq & \ \|d_k\|^2 - \|A_k d_k\|^2 / \|A_k\|^2 \end{aligned}$$

and so

$$\theta \left(\|d_k\|^2 - \|A_k d_k\|^2 / \|A_k\|^2 \right) \leq d_k^T W_k d_k$$

implies

$$\theta \|u_k\|^2 \leq d_k^T W_k d_k$$

Preliminary Algorithm

Apply an iterative solver to the primal-dual system

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

If

$$\begin{aligned} \theta \|d_k\|^2 &> \|A_k d_k\|^2 / \|A_k\|^2 \\ \text{and} \quad \theta \left(\|d_k\|^2 - \|A_k d_k\|^2 / \|A_k\|^2 \right) &> d_k^T W_k d_k \end{aligned}$$

then set $W_k \leftarrow \tilde{W}_k$ such that

$$\theta \left(\|d_k\|^2 - \|A_k d_k\|^2 / \|A_k\|^2 \right) \leq d_k^T \tilde{W}_k d_k$$

and continue the iteration

Outline

Problem Formulation

Unconstrained nonlinear optimization

$$\min_{x \in \boldsymbol{R}^n} \; f(x)$$

SQP subproblem

$$\min_{d \in \boldsymbol{R}^n} \; f_k + g_k^T d + \frac{1}{2} d^T H_k d$$

Step can be computed via Newton system

$$H_k d_k = -g_k$$

Problem Formulation

Unconstrained nonlinear optimization

$$\min_{x \in \boldsymbol{R}^n} \; f(x)$$

SQP subproblem

$$\min_{d \in \boldsymbol{R}^n} \; f_k + g_k^T d + \tfrac{1}{2} d^T H_k d$$

Step can be computed via Newton system

$$H_k d_k = -g_k$$

Problem Formulation

Unconstrained nonlinear optimization

$$\min_{x \in \mathbf{R}^n} \; f(x)$$

SQP subproblem

$$\min_{d \in \mathbf{R}^n} \; f_k + g_k^T d + \frac{1}{2} d^T H_k d$$

Step can be computed via Newton system

$$H_k d_k = -g_k$$

Applying an iterative solver

Inexact step satisfies

$$H_k d_k = -g_k + \rho_k$$

Line search method converges if the angle condition

$$\frac{-g_k^T d_k}{\|g_k\|\|d_k\|} \geq \theta$$

is satisfied, which can be verified directly

Applying an iterative solver

Inexact step satisfies

$$H_k d_k = -g_k + \rho_k$$

Line search method converges if the angle condition

$$\frac{-g_k^T d_k}{\|g_k\| \|d_k\|} \geq \theta$$

is satisfied, which can be verified directly

Outline

Hypothetical situation

Suppose we cannot verify an angle condition directly

$$g_k^T d_k \leq -\theta \|g_k\| \|d_k\|$$

- Verified indirectly if

$$d_k^T W_k d_k \geq \theta \|d_k\|^2 \quad \text{and} \quad \|\rho_k\| \leq \theta \|g_k\|$$

($W_k \leftarrow \tilde{W}_k$ in a line search method)

- Verified indirectly if

$$g_k^T d_k \leq -\theta \|d_k\|^2 \quad \text{and} \quad \|d_k\| \geq \theta \|g_k\|$$

(Δ_k reduced in a trust region method)

Hypothetical situation

Suppose we cannot verify an angle condition directly

$$g_k^T d_k \leq -\theta \|g_k\| \|d_k\|$$

- Verified indirectly if

$$d_k^T W_k d_k \geq \theta \|d_k\|^2 \quad \text{and} \quad \|\rho_k\| \leq \theta \|g_k\|$$

($W_k \leftarrow \tilde{W}_k$ in a line search method)

- Verified indirectly if

$$g_k^T d_k \leq -\theta \|d_k\|^2 \quad \text{and} \quad \|d_k\| \geq \theta \|g_k\|$$

(Δ_k reduced in a trust region method)

Hypothetical situation

Suppose we cannot verify an angle condition directly

$$g_k^T d_k \leq -\theta \|g_k\| \|d_k\|$$

- Verified indirectly if

$$d_k^T W_k d_k \geq \theta \|d_k\|^2 \quad \text{and} \quad \|\rho_k\| \leq \theta \|g_k\|$$

($W_k \leftarrow \tilde{W}_k$ in a line search method)

- Verified indirectly if

$$g_k^T d_k \leq -\theta \|d_k\|^2 \quad \text{and} \quad \|d_k\| \geq \theta \|g_k\|$$

(Δ_k reduced in a trust region method)

Outline

Inexact SQP Framework

Recall step computation

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

Step quality ensured by Stopping Conditions I and II

- SC I: descent for ϕ_π for current π
- SC II: descent for ϕ_π for increased π

Step length ensured by problem characteristics

- ... but properties are lost for indefinite $Z_k^T W_k Z_k$

Inexact SQP Framework

Recall step computation

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

Step quality ensured by Stopping Conditions I and II

- SC I: descent for ϕ_π for current π
- SC II: descent for ϕ_π for increased π

Step length ensured by problem characteristics

- … but properties are lost for indefinite $Z_k^T W_k Z_k$

Inexact SQP Framework

Recall step computation

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

Step quality ensured by Stopping Conditions I and II

- SC I: descent for ϕ_π for current π
- SC II: descent for ϕ_π for increased π

Step length ensured by problem characteristics

- ... but properties are lost for indefinite $Z_k^T W_k Z_k$

Inexact SQP Framework

Recall step computation

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

Step quality ensured by Stopping Conditions I and II

- SC I: descent for ϕ_π for current π
- SC II: descent for ϕ_π for increased π

Step length ensured by problem characteristics

- ... but properties are lost for indefinite $Z_k^T W_k Z_k$

"Angle Test" for Constrained Optimization

An inexact SQP step is acceptable if it satisfies an "angle test" for the penalty function ϕ_π

- Taylor expansion yields

$$D\phi_\pi(d_k) \leq g_k^T d_k - \pi(\|c_k\| - \|r_k\|)$$

- What is the steepest descent direction? (π not fixed)

"Angle test" cannot be verified directly

Outline

Sufficient Descent

Step satisfying Stopping Condition I or II is acceptable if

$$\begin{aligned} D\phi_\pi(d_k) &\leq -\theta\|d_k\|^2 \\ \text{and} \quad D\phi_\pi(d_k) &\leq -\theta\left(\|u_k\|^2 + \|c_k\|\right) \end{aligned}$$

which hold if

$$\begin{aligned} D\phi_\pi(d_k) &\leq -\theta\|d_k\|^2 \\ \text{and} \quad D\phi_\pi(d_k) &\leq -\theta\left(\|d_k\|^2 - \|A_k d_k\|^2/\|A_k\|^2 + \|c_k\|\right) \end{aligned}$$

where π is current (SC I) or increased (bounded) value (SC II)

Proposed Algorithm

Apply an iterative solver to the primal-dual system until SC I or II is satisfied

- Accept step if
  - ... tangential step is bounded

$$\theta\|d_k\|^2 \leq \|A_k d_k\|^2/\|A_k\|^2$$

  - ... or curvature is sufficiently positive

$$\theta\left(\|d_k\|^2 - \|A_k d_k\|^2/\|A_k\|^2\right) \leq d_k^T W_k d_k$$

  - ... or direction is of sufficient descent

$$\begin{aligned} D\phi_\pi(d_k) \leq & -\theta\|d_k\|^2 \\ \text{and} \quad D\phi_\pi(d_k) \leq & -\theta\left(\|d_k\|^2 - \|A_k d_k\|^2/\|A_k\|^2 + \|c_k\|\right) \end{aligned}$$

- If $\theta\|d_k\|^2 > \|A_k d_k\|^2/\|A_k\|^2$, then set $W_k \leftarrow \tilde{W}_k$ to satisfy

$$\theta\left(\|d_k\|^2 - \|A_k d_k\|^2/\|A_k\|^2\right) \leq d_k^T \tilde{W}_k d_k$$

Conclusion

We have

- ... observed that an inexact line search SQP algorithm may fail if negative curvature is present
- ... proposed a method for modifying W_k during the application of the iterative solver to ensure global convergence
- ... proposed techniques for avoiding modifications of W_k

Final Note on Modifying W_k

The Hessian of the Lagrangian has the form

$$W_k = \nabla^2_{xx} f(x) + \sum_{i=1}^{m} \lambda^i \nabla^2_{xx} c^i(x)$$

so we may consider modifications of the form

$$\tilde{W}_k = \nabla^2_{xx} f(x) + \sum_{i=1}^{m} \lambda^i \nabla^2_{xx} c^i(x) + G_k$$

where $G_k \succeq 0$, or

$$\tilde{W}_k = \nabla^2_{xx} f(x) + \sum_{i \in \mathcal{I}} \lambda^i \nabla^2_{xx} c^i(x)$$

if $\nabla^2_{xx} f(x) \succ 0$ is known and $\mathcal{I} \subseteq \{1, \ldots, m\}$

Thanks!