Derivative Free Optimization of Nonlinear Functions

Some New Results

Andrew R. Conn
arconn@us.ibm.com

Mathematical Sciences
IBM T.J. Watson Research Center

with special thanks to my coauthors Luís Nunes Vicente and Katya Scheinberg

Eighth US-Mexico Workshop on Optimization and its Applications, 2007, Huatulco

Basic Algorithm

Initialize: x_0, Δ

Compute Model: $m_k(\ )$

Compute Step: Compute s_k from

$$\min_{\|s\| \leq \Delta} m_k(x_k + s)$$

Trust-region Update: $\rho = \frac{f(x_k) - f(x_k + s_k)}{m_k(x_k) - m_k(x_k + s_k)}$

If $\rho > 0.75$ $\Delta \leftarrow 2.0\Delta$ Accept $x_k + s_k$

If $0.25 < \rho < 0.75$ $\Delta \leftarrow \Delta$ Accept $x_k + s_k$

If $\rho < 0.25$ $\Delta \leftarrow 0.5\Delta$ Reject $x_k + s_k$

Basic Algorithm

Initialize: x_0, Δ

Compute Model: $m_k(\ )$

Compute Step: Compute s_k from

$$\min_{\|s\| \leq \Delta} m_k(x_k + s)$$

Trust-region Update: $\rho = \frac{f(x_k) - f(x_k + s_k)}{m_k(x_k) - m_k(x_k + s_k)}$

If $\rho > 0.75$ $\Delta \leftarrow 2.0\Delta$ Accept $x_k + s_k$

If $0.25 < \rho < 0.75$ $\Delta \leftarrow \Delta$ Accept $x_k + s_k$

If $\rho < 0.25$ $\Delta \leftarrow 0.5\Delta$ Reject $x_k + s_k$

Basic Algorithm

Initialize: x_0, Δ

Compute Model: $m_k(\ )$

Compute Step: Compute s_k from

$$\min_{\|s\| \leq \Delta} m_k(x_k + s)$$

Trust-region Update: $\rho = \frac{f(x_k) - f(x_k + s_k)}{m_k(x_k) - m_k(x_k + s_k)}$

If $\rho > 0.75$ $\quad \Delta \leftarrow 2.0\Delta$ $\quad$ Accept $x_k + s_k$

If $0.25 < \rho < 0.75$ $\quad \Delta \leftarrow \Delta$ $\quad$ Accept $x_k + s_k$

If $\rho < 0.25$ $\quad \Delta \leftarrow 0.5\Delta$ $\quad$ Reject $x_k + s_k$

Basic Algorithm

Initialize: x_0, Δ

Compute Model: $m_k(\ )$

Compute Step: Compute s_k from

$$\min_{\|s\| \leq \Delta} m_k(x_k + s)$$

Trust-region Update: $\rho = \frac{f(x_k) - f(x_k + s_k)}{m_k(x_k) - m_k(x_k + s_k)}$

If $\rho > 0.75$ $\quad \Delta \leftarrow 2.0\Delta$ $\quad$ Accept $x_k + s_k$

If $0.25 < \rho < 0.75$ $\quad \Delta \leftarrow \Delta$ $\quad$ Accept $x_k + s_k$

If $\rho < 0.25$ $\quad \Delta \leftarrow 0.5\Delta$ $\quad$ Reject $x_k + s_k$

Basic Algorithm

Initialize: x_0, Δ

Compute Model: $m_k(\ )$

Compute Step: Compute s_k from

$$\min_{\|s\| \leq \Delta} m_k(x_k + s)$$

Trust-region Update: $\rho = \frac{f(x_k) - f(x_k + s_k)}{m_k(x_k) - m_k(x_k + s_k)}$

If $\rho > 0.75$ $\Delta \leftarrow 2.0\Delta$ Accept $x_k + s_k$

If $0.25 < \rho < 0.75$ $\Delta \leftarrow \Delta$ Accept $x_k + s_k$

If $\rho < 0.25$ $\Delta \leftarrow 0.5\Delta$ Reject $x_k + s_k$

Points to note

- Model depends on previous iterates!
- Geometry matters

- In derivative free methods we use sample based models; e.g., interpolation or regression.
- The $\mathcal{O}$ in Taylor-like bounds depends not only on f, but also on the geometry of the sample set.
- We need to have some constant characterizing the quality of the sample set.
- We need to control this constant to keep it uniformly bounded.
- A Fully Linear model that is suitably minimized replaces the Cauchy Point.

Points to note

- Model depends on previous iterates!
- Geometry matters
- In derivative free methods we use sample based models; e.g., interpolation or regression.
- The $\mathcal{O}$ in Taylor-like bounds depends not only on f, but also on the geometry of the sample set.
- We need to have some constant characterizing the quality of the sample set.
- We need to control this constant to keep it uniformly bounded.
- A Fully Linear model that is suitably minimized replaces the Cauchy Point.

Points to note

- Model depends on previous iterates!
- Geometry matters

- In derivative free methods we use sample based models; e.g., interpolation or regression.
- The $\mathcal{O}$ in Taylor-like bounds depends not only on f, but also on the geometry of the sample set.
- We need to have some constant characterizing the quality of the sample set.
- We need to control this constant to keep it uniformly bounded.
- A Fully Linear model that is suitably minimized replaces the Cauchy Point.

Points to note

- Model depends on previous iterates!
- Geometry matters
- In derivative free methods we use sample based models; e.g., interpolation or regression.
- The $\mathcal{O}$ in Taylor-like bounds depends not only on f, but also on the geometry of the sample set.
- We need to have some constant characterizing the quality of the sample set.
- We need to control this constant to keep it uniformly bounded.
- A Fully Linear model that is suitably minimized replaces the Cauchy Point.

Points to note

- Model depends on previous iterates!
- Geometry matters

- In derivative free methods we use sample based models; e.g., interpolation or regression.
- The $\mathcal{O}$ in Taylor-like bounds depends not only on f, but also on the geometry of the sample set.
- We need to have some constant characterizing the quality of the sample set.
- We need to control this constant to keep it uniformly bounded.
- A Fully Linear model that is suitably minimized replaces the Cauchy Point.

Abstraction of required bounds

- $f \in C^1$ and ∇f Lipschitz continuous on $\{x | f_k \leq f_0\}$.
- Δ_k bounded above.

Abstraction of required bounds

- $f \in C^1$ and ∇f Lipschitz continuous on $\{x | f_k \leq f_0\}$.
- Δ_k bounded above.

Abstraction of required bounds

- $f \in C^1$ and ∇f Lipschitz continuous on $\{x | f_k \leq f_0\}$.
- Δ_k bounded above.
- An interpolation model using at least $n + 1$ poised points that are contained in $B(x, \Delta)$ and satisfies Taylor-like error bounds is called: Fully Linear on $B(x, \Delta)$ iff

$$|f(x) - m(x)| \leq \kappa_{ef} \Delta^2$$

$$|\nabla f(x) - \nabla m(x)| \leq \kappa_{eg} \Delta$$

for all x in $B(x, \Delta)$

Abstraction of required bounds

- $f \in C^1$ and ∇f Lipschitz continuous on $\{x | f_k \leq f_0\}$.
- Δ_k bounded above.
- A model is called: Fully Linear on $B(x, \Delta)$ iff

$$|f(x) - m(x)| \leq \kappa_{ef} \Delta^2$$

$$|\nabla f(x) - \nabla m(x)| \leq \kappa_{eg} \Delta$$

for all x in $B(x, \Delta)$

Abstraction of required bounds

- $f \in C^1$ and ∇f Lipschitz continuous on $\{x | f_k \leq f_0\}$.
- $f \in C^2$ and $\nabla^2 f$ Lipschitz continuous on $\{x | f_k \leq f_0\}$.
- Δ_k bounded above.
- A model is called: Fully Quadratic on $B(x, \Delta)$ iff

$$|f(x) - m(x)| \leq \kappa_{ef} \Delta^3$$

$$|\nabla f(x) - \nabla m(x)| \leq \kappa_{eg} \Delta^2$$

$$|\nabla^2 f(x) - \nabla^2 m(x)| \leq \kappa_{eh} \Delta$$

for all x in $B(x, \Delta)$

Interpolation-based trust-region methods

- Derivative-free methods.
- Trust region subproblems are defined by linear/quadratic models built by multivariate polynomial interpolation.
- Need to deal with approximated gradients and, when using quadratic models, with approximated Hessians.

- Trust region methods retain global convergence to stationarity if:
  - interpolation models are at least fully linear;
  - well-poisedness is enforced before reducing the trust radius;
  - well-poisedness is enforced once the size of the model gradient becomes small.

- Global convergence to second order critical points possible when using fully quadratic models.

Interpolation-based trust-region methods

- Derivative-free methods.
- Trust region subproblems are defined by linear/quadratic models built by multivariate polynomial interpolation.
- Need to deal with approximated gradients and, when using quadratic models, with approximated Hessians.

- Trust region methods retain global convergence to stationarity if:
  - interpolation models are at least fully linear;
  - well-poisedness is enforced before reducing the trust radius;
  - well-poisedness is enforced once the size of the model gradient becomes small.

- Global convergence to second order critical points possible when using fully quadratic models.

Interpolation-based trust-region methods

- Derivative-free methods.
- Trust region subproblems are defined by linear/quadratic models built by multivariate polynomial interpolation.
- Need to deal with approximated gradients and, when using quadratic models, with approximated Hessians.

- Trust region methods retain global convergence to stationarity if:
  - interpolation models are at least fully linear;
  - well-poisedness is enforced before reducing the trust radius;
  - well-poisedness is enforced once the size of the model gradient becomes small.

- Global convergence to second order critical points possible when using fully quadratic models.

Interpolation-based trust-region methods

- Derivative-free methods.
- Trust region subproblems are defined by linear/quadratic models built by multivariate polynomial interpolation.
- Need to deal with approximated gradients and, when using quadratic models, with approximated Hessians.

- Trust region methods retain global convergence to stationarity if:
  - interpolation models are at least fully linear;
  - well-poisedness is enforced before reducing the trust radius;
  - well-poisedness is enforced once the size of the model gradient becomes small.

- Global convergence to second order critical points possible when using fully quadratic models.

Interpolation-based trust-region methods

- Derivative-free methods.
- Trust region subproblems are defined by linear/quadratic models built by multivariate polynomial interpolation.
- Need to deal with approximated gradients and, when using quadratic models, with approximated Hessians.

- Trust region methods retain global convergence to stationarity if:
  - interpolation models are at least fully linear;
  - well-poisedness is enforced before reducing the trust radius;
  - well-poisedness is enforced once the size of the model gradient becomes small.

- Global convergence to second order critical points possible when using fully quadratic models.

Interpolation-based trust-region methods

- Derivative-free methods.
- Trust region subproblems are defined by linear/quadratic models built by multivariate polynomial interpolation.
- Need to deal with approximated gradients and, when using quadratic models, with approximated Hessians.

- Trust region methods retain global convergence to stationarity if:
  - interpolation models are at least fully linear;
  - well-poisedness is enforced before reducing the trust radius;
  - well-poisedness is enforced once the size of the model gradient becomes small.

- Global convergence to second order critical points possible when using fully quadratic models.

Interpolation-based trust-region methods

- Derivative-free methods.
- Trust region subproblems are defined by linear/quadratic models built by multivariate polynomial interpolation.
- Need to deal with approximated gradients and, when using quadratic models, with approximated Hessians.

- Trust region methods retain global convergence to stationarity if:
  - interpolation models are at least fully linear;
  - well-poisedness is enforced before reducing the trust radius;
  - well-poisedness is enforced once the size of the model gradient becomes small.

- Global convergence to second order critical points possible when using fully quadratic models.

Interpolation-based trust-region methods

- Derivative-free methods.
- Trust region subproblems are defined by linear/quadratic models built by multivariate polynomial interpolation.
- Need to deal with approximated gradients and, when using quadratic models, with approximated Hessians.

- Trust region methods retain global convergence to stationarity if:
    - interpolation models are at least fully linear;
    - well-poisedness is enforced before reducing the trust radius;
    - well-poisedness is enforced once the size of the model gradient becomes small.

- Global convergence to second order critical points possible when using fully quadratic models.

Model decrease at the Cauchy point+Eigenpoint

Fundamental result:

$$m_k(x_k) - m_k(x_k + s_k) \geq \frac{\kappa}{2} \max \left\{ \|g_k\| \min \left[\frac{\|g_k\|}{\kappa_{bhd}}, \Delta_k \right], -\tau_k \Delta_k^2 \right\}.$$

where

$$g_k = \nabla_x m_k(x_k), \quad \beta_k = 1 + \max_{x \in \mathcal{B}_k} \|\nabla_{xx} m_k(x)\|$$

,

$$x_k^C(t) = \{x \mid x = x_k - t g_k, \; t \geq 0 \text{ and } x \in \mathcal{B}_k\}.$$

and τ is the most negative eigenvalue

Model decrease at the Cauchy point+Eigenpoint

Fundamental result:

$$m_k(x_k) - m_k(x_k + s_k) \geq \frac{\kappa}{2} \max\left\{ \|g_k\| \min\left[\frac{\|g_k\|}{\kappa_{bhd}}, \Delta_k\right], -\tau_k \Delta_k^2 \right\}.$$

where

$$g_k = \nabla_x m_k(x_k), \quad \beta_k = 1 + \max_{x \in \mathcal{B}_k} \|\nabla_{xx} m_k(x)\|$$

,

$$x_k^C(t) = \{x \mid x = x_k - t g_k,\ t \geq 0 \ \text{ and } \ x \in \mathcal{B}_k\}.$$

and τ is the most negative eigenvalue

Model decrease at the Cauchy point+Eigenpoint

Fundamental result:

$$m_k(x_k) - m_k(x_k + s_k) \geq \frac{\kappa}{2} \max\left\{ \|g_k\| \min\left[\frac{\|g_k\|}{\kappa_{bhd}}, \Delta_k\right], -\tau_k \Delta_k^2 \right\}.$$

where

$$g_k = \nabla_x m_k(x_k), \quad \beta_k = 1 + \max_{x \in \mathcal{B}_k} \|\nabla_{xx} m_k(x)\|$$

,

$$x_k^{\mathrm{C}}(t) = \{x \mid x = x_k - t g_k, \; t \geq 0 \;\text{ and }\; x \in \mathcal{B}_k\}.$$

and τ is the most negative eigenvalue

Global convergence to 2^{nd}-order critical points

Need

- $\nabla^2 f$ Lipshitz continuous.
- Fully-quadratic (FQ) models.
  i.e. satisfies the correct Taylor-like bounds

We know we can achieve FQ in **finite number of steps.**

Need to define stationarity criteria

- $\sigma_k^m = \|g_k\| + \max\{-\lambda_{\min}(\nabla^2 m_k), 0\}$
- $\sigma_k = \|\nabla f_k\| + \max\{-\lambda_{\min}(\nabla^2 f_k), 0\}$

Criticality Test: If $\|\sigma_k^m\| \leq \epsilon_c$, construct a model, $\tilde{m}_k$ that is FQ on $B_k(x_k, \Delta_k)$, for some $\tilde{\Delta}_k \in (0, \mu_k \sigma_k^m)$, using the algorithm on the next slide

Set $m_k = \tilde{m}_k$ and $\Delta_k = \min\{\tilde{\Delta}_k, \Delta_k\}$.

Global convergence to 2^{nd}-order critical points

Need

- $\nabla^2 f$ Lipshitz continuous.
- Fully-quadratic (FQ) models.
  i.e. satisfies the correct Taylor-like bounds

We know we can achieve FQ in **finite number of steps.**

Need to define stationarity criteria

- $\sigma_k^m = \|g_k\| + \max\{-\lambda_{\min}(\nabla^2 m_k), 0\}$
- $\sigma_k = \|\nabla f_k\| + \max\{-\lambda_{\min}(\nabla^2 f_k), 0\}$

Criticality Test: If $\|\sigma_k^m\| \leq \epsilon_c$, construct a model, $\tilde{m}_k$ that is FQ on $B_k(x_k, \Delta_k)$, for some $\tilde{\Delta}_k \in (0, \mu_k \sigma_k^m)$, using the algorithm on the next slide

Set $m_k = \tilde{m}_k$ and $\Delta_k = \min\{\tilde{\Delta}_k, \Delta_k\}$.

Global convergence to 2^{nd}-order critical points

Need

- $\nabla^2 f$ Lipshitz continuous.
- Fully-quadratic (FQ) models.
  i.e. satisfies the correct Taylor-like bounds

We know we can achieve FQ in **finite number of steps.**

Need to define stationarity criteria

- $\sigma_k^m = \|g_k\| + \max\{-\lambda_{\min}(\nabla^2 m_k), 0\}$
- $\sigma_k = \|\nabla f_k\| + \max\{-\lambda_{\min}(\nabla^2 f_k), 0\}$

Criticality Test: If $\|\sigma_k^m\| \leq \epsilon_c$, construct a model, $\tilde{m}_k$ that is FQ on $B_k(x_k, \Delta_k)$, for some $\tilde{\Delta}_k \in (0, \mu_k \sigma_k^m)$, using the algorithm on the next slide

Set $m_k = \tilde{m}_k$ and $\Delta_k = \min\{\tilde{\Delta}_k, \Delta_k\}$.

Global convergence to 2^{nd}-order critical points

Need

- $\nabla^2 f$ Lipshitz continuous.
- Fully-quadratic (FQ) models.
  i.e. satisfies the correct Taylor-like bounds

We know we can achieve FQ in **finite number of steps.**

Need to define stationarity criteria

- $\sigma_k^m = \|g_k\| + \max\{-\lambda_{\min}(\nabla^2 m_k), 0\}$
- $\sigma_k = \|\nabla f_k\| + \max\{-\lambda_{\min}(\nabla^2 f_k), 0\}$

Criticality Test: If $\|\sigma_k^m\| \leq \epsilon_c$, construct a model, $\tilde{m}_k$ that is FQ on $B_k(x_k, \Delta_k)$, for some $\tilde{\Delta}_k \in (0, \mu_k \sigma_k^m)$, using the algorithm on the next slide

Set $m_k = \tilde{m}_k$ and $\Delta_k = \min\{\tilde{\Delta}_k, \Delta_k\}$.

Global convergence to 2^{nd}-order critical points

Need

- $\nabla^2 f$ Lipshitz continuous.
- Fully-quadratic (FQ) models.
  i.e. satisfies the correct Taylor-like bounds

We know we can achieve FQ in **finite number of steps.**

Need to define stationarity criteria

- $\sigma_k^m = \|g_k\| + \max\{-\lambda_{\min}(\nabla^2 m_k), 0\}$
- $\sigma_k = \|\nabla f_k\| + \max\{-\lambda_{\min}(\nabla^2 f_k), 0\}$

Criticality Test: If $\|\sigma_k^m\| \leq \epsilon_c$, construct a model, $\tilde{m}_k$ that is FQ on $B_k(x_k, \Delta_k)$, for some $\tilde{\Delta}_k \in (0, \mu_k \sigma_k^m)$, using the algorithm on the next slide

Set $m_k = \tilde{m}_k$ and $\Delta_k = \min\{\tilde{\Delta}_k, \Delta_k\}$.

Global convergence to 2^{nd}-order critical points

Need

- $\nabla^2 f$ Lipshitz continuous.
- Fully-quadratic (FQ) models.
  i.e. satisfies the correct Taylor-like bounds

We know we can achieve FQ in **finite number of steps.**

Need to define stationarity criteria

- $\sigma_k^m = \|g_k\| + \max\{-\lambda_{\min}(\nabla^2 m_k), 0\}$
- $\sigma_k = \|\nabla f_k\| + \max\{-\lambda_{\min}(\nabla^2 f_k), 0\}$

Criticality Test: If $\|\sigma_k^m\| \leq \epsilon_c$, construct a model, $\tilde{m}_k$ that is FQ on $B_k(x_k, \Delta_k)$, for some $\tilde{\Delta}_k \in (0, \mu_k \sigma_k^m)$, using the algorithm on the next slide

Set $m_k = \tilde{m}_k$ and $\Delta_k = \min\{\tilde{\Delta}_k, \Delta_k\}$.

Global convergence to 2^{nd}-order critical points

Need

- $\nabla^2 f$ Lipshitz continuous.
- Fully-quadratic (FQ) models.
  i.e. satisfies the correct Taylor-like bounds

We know we can achieve FQ in **finite number of steps.**

Need to define stationarity criteria

- $\sigma_k^m = \|g_k\| + \max\{-\lambda_{\min}(\nabla^2 m_k), 0\}$
- $\sigma_k = \|\nabla f_k\| + \max\{-\lambda_{\min}(\nabla^2 f_k), 0\}$

Criticality Test: If $\|\sigma_k^m\| \leq \epsilon_c$, construct a model, $\tilde{m}_k$ that is FQ on $B_k(x_k, \Delta_k)$, for some $\tilde{\Delta}_k \in (0, \mu_k \sigma_k^m)$, using the algorithm on the next slide

Set $m_k = \tilde{m}_k$ and $\Delta_k = \min\{\tilde{\Delta}_k, \Delta_k\}$.

Global convergence to 2^{nd}-order critical points

Need

- $\nabla^2 f$ Lipshitz continuous.
- Fully-quadratic (FQ) models.
  i.e. satisfies the correct Taylor-like bounds

We know we can achieve FQ in **finite number of steps.**

Need to define stationarity criteria

- $\sigma_k^m = \|g_k\| + \max\{-\lambda_{\min}(\nabla^2 m_k), 0\}$
- $\sigma_k = \|\nabla f_k\| + \max\{-\lambda_{\min}(\nabla^2 f_k), 0\}$

Criticality Test: If $\|\sigma_k^m\| \leq \epsilon_c$, construct a model, $\tilde{m}_k$ that is FQ on $B_k(x_k, \Delta_k)$, for some $\tilde{\Delta}_k \in (0, \mu_k \sigma_k^m)$, using the algorithm on the next slide

Set $m_k = \tilde{m}_k$ and $\Delta_k = \min\{\tilde{\Delta}_k, \Delta_k\}$.

2^{nd}-Order Criticality Step Algorithm (CSA)

This algorithm is only applied if the model m_k is not FQ or if $\Delta_k > \mu\sigma_k^m$.

[Initialization:] Set $i = 0$. Choose $\alpha \in (0, 1)$. Set $m_k^{(0)} = m_k$.

[Repeat] $i \leftarrow i + 1$. Improve $m_k^{(i-1)}$ until it is fully quadratic on $B(x_k; \alpha^i \mu \|(\sigma_k^m)^{(0)}\|)$ (notice that this can be done in a finite, uniformly bounded number of steps). Denote the new model by $m_k^{(i)}$.

Set $\tilde{\Delta}_k = \alpha^i \mu \|(\sigma_k^m)^{(0)}\|$ and $\tilde{m}_k = m_k^{(i)}$.

[Until] $\tilde{\Delta}_k \leq \mu \|(\sigma_k^m)^{(i)}\|$.

Ensures that close to stationary ⇒ FQ on a related TR ⇒ true gradient small

2^{nd}-Order Criticality Step Algorithm (CSA)

This algorithm is only applied if the model m_k is not FQ or if $\Delta_k > \mu\sigma_k^m$.

[Initialization:] Set $i = 0$. Choose $\alpha \in (0, 1)$. Set $m_k^{(0)} = m_k$.

[Repeat] $i \leftarrow i + 1$. Improve $m_k^{(i-1)}$ until it is fully quadratic on $B(x_k; \alpha^i \mu \|(\sigma_k^m)^{(0)}\|)$ (notice that this can be done in a finite, uniformly bounded number of steps). Denote the new model by $m_k^{(i)}$.

Set $\tilde{\Delta}_k = \alpha^i \mu \|(\sigma_k^m)^{(0)}\|$ and $\tilde{m}_k = m_k^{(i)}$.

[Until] $\tilde{\Delta}_k \leq \mu \|(\sigma_k^m)^{(i)}\|$.

Ensures that close to stationary ⇒ FQ on a related TR ⇒ true gradient small

2^{nd}-Order Criticality Step Algorithm (CSA)

This algorithm is only applied if the model m_k is not FQ or if $\Delta_k > \mu\sigma_k^m$.

[Initialization:] Set $i = 0$. Choose $\alpha \in (0, 1)$. Set $m_k^{(0)} = m_k$.

[Repeat] $i \leftarrow i + 1$. Improve $m_k^{(i-1)}$ until it is fully quadratic on $B(x_k; \alpha^i \mu \|(\sigma_k^m)^{(0)}\|)$ (notice that this can be done in a finite, uniformly bounded number of steps). Denote the new model by $m_k^{(i)}$.

Set $\tilde{\Delta}_k = \alpha^i \mu \|(\sigma_k^m)^{(0)}\|$ and $\tilde{m}_k = m_k^{(i)}$.

[Until] $\tilde{\Delta}_k \leq \mu \|(\sigma_k^m)^{(i)}\|$.

Ensures that close to stationary ⇒ FQ on a related TR ⇒ true gradient small

2^{nd}-Order Criticality Step Algorithm (CSA)

This algorithm is only applied if the model m_k is not FQ or if $\Delta_k > \mu \sigma_k^m$.

[Initialization:] Set $i = 0$. Choose $\alpha \in (0, 1)$. Set $m_k^{(0)} = m_k$.

[Repeat] $i \leftarrow i + 1$. Improve $m_k^{(i-1)}$ until it is fully quadratic on $B(x_k; \alpha^i \mu \|(\sigma_k^m)^{(0)}\|)$ (notice that this can be done in a finite, uniformly bounded number of steps). Denote the new model by $m_k^{(i)}$.

Set $\tilde{\Delta}_k = \alpha^i \mu \|(\sigma_k^m)^{(0)}\|$ and $\tilde{m}_k = m_k^{(i)}$.

[Until] $\tilde{\Delta}_k \leq \mu \|(\sigma_k^m)^{(i)}\|$.

Ensures that close to stationary ⇒ FQ on a related TR ⇒ true gradient small

The Algorithm –2^{nd} order version

Step 0: Initialization. x_0, $\Delta_{\max}$, $\Delta_0 \in (0, \Delta_{\max})$, and a initial model m_0, β, $\epsilon_c > 0$ and $\mu > \beta > 0$, $\eta_1 \neq 0$, η_0 and γ with $0 \leq \eta_0 \leq \eta_1 < 1$ and $0 < \gamma < 1 < \gamma_+$. Set $k = 0$.

Step 1: Criticality test. If $\sigma_k^m \leq \epsilon_c$, use CSA to construct $\tilde{m}_k$, FQ (for κ_{ef}, κ_{eg}, and κ_{eh}, same for all iterations) on $B(x_k; \tilde{\Delta}_k)$ for some $\tilde{\Delta}_k \in (0, \mu \|\tilde{\sigma}_k^m\|]$. Set $m_k = \tilde{m}_k$ and $\Delta_k = \min\{\tilde{\Delta}_k, \Delta_k\}$.

Step 2: Step calculation. Choose a step s_k that "sufficiently reduces the model" m_k (approximate CP/EP) such that $x_k + s_k \in \mathcal{B}_k(x_k; \Delta_k)$.

The Algorithm (continued)

Step 3: Acceptance of the trial point. Compute $f(x_k + s_k)$ and
$\rho_k = \frac{f(x_k) - f(x_k + s_k)}{m_k(x_k) - m_k(x_k + s_k)}$.
If $\rho_k > \eta_1$ or $\rho_k > \eta_0$ and m_k is FQ on $B(x_k; \Delta_k)$,
then $x_{k+1} = x_k + s_k$ and the model is updated;
otherwise the model and the iterate remain
unchanged.

Step 4: Model improvement. If $\rho_k < \eta_1$ ensure m_k is FQ, making
suitable improvements if necessary and define m_{k+1}
to be the (possibly improved) model.

The Algorithm (continued)

Step 5: Trust-region radius update. Set

$$\Delta_{k+1} \in \begin{cases} \gamma_+ \Delta_k & \text{if } \rho_k \geq \eta_1, m_k \text{ is FQ, and } \Delta_k < \beta \sigma_k^m \\ [\Delta_k, \Delta_{\max}] & \text{if } \rho_k \geq \eta_1, m_k \text{ is FQ, and } \Delta_k \geq \beta \sigma_k^m \\ \gamma \Delta_k & \text{if } \rho_k < \eta_1, \text{ and } m_k \text{ is FQ,} \\ \Delta_k & \text{if } m_k \text{ is not FQ.} \end{cases}$$

Increment k by one and go to Step 1.

Step 1 Outcome

1. $\rho_k \geq \eta_1$ **Successful** iteration. $\Delta_{k+1} \geq \Delta_k$

   2 kinds if $\eta_0 = 0$: Only increase Δ when Δ small relative to criticality!

2. $\eta_1 > \rho_k \geq \eta_0$ and m_k is FQ. **Acceptable** iteration. $\Delta_{k+1} < \Delta_k$

   NB There are no acceptable iterations if $\eta_0 = \eta_1 \in (0, 1)$

3. $\eta_1 > \rho_k$ and m_k is not FQ. The model must be improved. **model-improving** iteration. So Δ_k and x_k **not changed**

4. $\rho_k < \eta_0$ and m_k is FQ. $\Delta_{k+1} < \Delta_k$ and $x_{k+1} = x_k$ **Unsuccessful** iteration.

Step 1 Outcome

1. $\rho_k \geq \eta_1$ **Successful** iteration. $\Delta_{k+1} \geq \Delta_k$

   2 kinds if $\eta_0 = 0$: Only increase Δ when Δ small relative to criticality!

2. $\eta_1 > \rho_k \geq \eta_0$ and m_k is FQ. **Acceptable** iteration. $\Delta_{k+1} < \Delta_k$

   NB There are no acceptable iterations if $\eta_0 = \eta_1 \in (0, 1)$

3. $\eta_1 > \rho_k$ and m_k is not FQ. The model must be improved. **model-improving** iteration. So Δ_k and x_k **not changed**

4. $\rho_k < \eta_0$ and m_k is FQ. $\Delta_{k+1} < \Delta_k$ and $x_{k+1} = x_k$ **Unsuccessful** iteration.

Step 1 Outcome

1. $\rho_k \geq \eta_1$ **Successful** iteration. $\Delta_{k+1} \geq \Delta_k$

   2 kinds if $\eta_0 = 0$: Only increase Δ when Δ small relative to criticality!

2. $\eta_1 > \rho_k \geq \eta_0$ and m_k is FQ. **Acceptable** iteration.
   $\Delta_{k+1} < \Delta_k$

   NB There are no acceptable iterations if $\eta_0 = \eta_1 \in (0, 1)$

3. $\eta_1 > \rho_k$ and m_k is not FQ. The model must be improved. **model-improving** iteration. So Δ_k and x_k **not changed**

4. $\rho_k < \eta_0$ and m_k is FQ. $\Delta_{k+1} < \Delta_k$ and $x_{k+1} = x_k$
   Unsuccessful iteration.

Step 1 Outcome

1. $\rho_k \geq \eta_1$ **Successful** iteration. $\Delta_{k+1} \geq \Delta_k$

   2 kinds if $\eta_0 = 0$: Only increase Δ when Δ small relative to criticality!

2. $\eta_1 > \rho_k \geq \eta_0$ and m_k is FQ. **Acceptable** iteration. $\Delta_{k+1} < \Delta_k$

   NB There are no acceptable iterations if $\eta_0 = \eta_1 \in (0, 1)$

3. $\eta_1 > \rho_k$ and m_k is not FQ. The model must be improved. **model-improving** iteration. So Δ_k and x_k **not changed**

4. $\rho_k < \eta_0$ and m_k is FQ. $\Delta_{k+1} < \Delta_k$ and $x_{k+1} = x_k$ **Unsuccessful** iteration.

Step 1 Outcome

1. $\rho_k \geq \eta_1$ **Successful** iteration. $\Delta_{k+1} \geq \Delta_k$

   2 kinds if $\eta_0 = 0$: Only increase Δ when Δ small relative to criticality!

2. $\eta_1 > \rho_k \geq \eta_0$ and m_k is FQ. **Acceptable** iteration. $\Delta_{k+1} < \Delta_k$

   NB There are no acceptable iterations if $\eta_0 = \eta_1 \in (0, 1)$

3. $\eta_1 > \rho_k$ and m_k is not FQ. The model must be improved. **model-improving** iteration. So Δ_k and x_k **not changed**

4. $\rho_k < \eta_0$ and m_k is FQ. $\Delta_{k+1} < \Delta_k$ and $x_{k+1} = x_k$ **Unsuccessful** iteration.

Convergence Highlights

- Convergence and not liminf convergence even if $\eta_0 = 0$ — as long as in the latter case we are careful when we increase Δ_k.
- If $\rho_k < \eta_1$ only reduce Δ_k if FQ.
- Criticality test includes curvature information
- Relatively general framework — which we can satisfy in a finite number of steps
- Do not require FQ/FL at every step
- Four possible outcomes: Successful, Acceptable, Model Improving, and Unsuccesful.
- After criticality step $\Delta_k = \min\{\tilde{\Delta}_k, \Delta_k\}$: good efficiency.
- All details of the proof are given

Convergence Highlights

- Convergence and not liminf convergence even if $\eta_0 = 0$ — as long as in the latter case we are careful when we increase Δ_k.
- If $\rho_k < \eta_1$ only reduce Δ_k if FQ.
- Criticality test includes curvature information
- Relatively general framework — which we can satisfy in a finite number of steps
- Do not require FQ/FL at every step
- Four possible outcomes: Successful, Acceptable, Model Improving, and Unsuccesful.
- After criticality step $\Delta_k = \min\{\tilde{\Delta}_k, \Delta_k\}$: good efficiency.
- All details of the proof are given

Convergence Highlights

- Convergence and not liminf convergence even if $\eta_0 = 0$ — as long as in the latter case we are careful when we increase Δ_k.
- If $\rho_k < \eta_1$ only reduce Δ_k if FQ.
- Criticality test includes curvature information
- Relatively general framework — which we can satisfy in a finite number of steps
- Do not require FQ/FL at every step
- Four possible outcomes: Successful, Acceptable, Model Improving, and Unsuccesful.
- After criticality step $\Delta_k = \min\{\tilde{\Delta}_k, \Delta_k\}$: good efficiency.
- All details of the proof are given

Convergence Highlights

- Convergence and not liminf convergence even if $\eta_0 = 0$ — as long as in the latter case we are careful when we increase Δ_k.
- If $\rho_k < \eta_1$ only reduce Δ_k if FQ.
- Criticality test includes curvature information
- Relatively general framework — which we can satisfy in a finite number of steps
- Do not require FQ/FL at every step
- Four possible outcomes: Successful, Acceptable, Model Improving, and Unsuccesful.
- After criticality step $\Delta_k = \min\{\tilde{\Delta}_k, \Delta_k\}$: good efficiency.
- All details of the proof are given

Convergence Highlights

- Convergence and not liminf convergence even if $\eta_0 = 0$ — as long as in the latter case we are careful when we increase Δ_k.
- If $\rho_k < \eta_1$ only reduce Δ_k if FQ.
- Criticality test includes curvature information
- Relatively general framework — which we can satisfy in a finite number of steps
- Do not require FQ/FL at every step
- Four possible outcomes: Successful, Acceptable, Model Improving, and Unsuccesful.
- After criticality step $\Delta_k = \min\{\tilde{\Delta}_k, \Delta_k\}$: good efficiency.
- All details of the proof are given

Convergence Highlights

- Convergence and not liminf convergence even if $\eta_0 = 0$ — as long as in the latter case we are careful when we increase Δ_k.
- If $\rho_k < \eta_1$ only reduce Δ_k if FQ.
- Criticality test includes curvature information
- Relatively general framework — which we can satisfy in a finite number of steps
- Do not require FQ/FL at every step
- Four possible outcomes: Successful, Acceptable, Model Improving, and Unsuccesful.
- After criticality step $\Delta_k = \min\{\tilde{\Delta}_k, \Delta_k\}$: good efficiency.
- All details of the proof are given

Convergence Highlights

- Convergence and not liminf convergence even if $\eta_0 = 0$ — as long as in the latter case we are careful when we increase Δ_k.
- If $\rho_k < \eta_1$ only reduce Δ_k if FQ.
- Criticality test includes curvature information
- Relatively general framework — which we can satisfy in a finite number of steps
- Do not require FQ/FL at every step
- Four possible outcomes: Successful, Acceptable, Model Improving, and Unsuccesful.
- After criticality step $\Delta_k = \min\{\tilde{\Delta}_k, \Delta_k\}$: good efficiency.
- All details of the proof are given

Convergence Highlights

- Convergence and not liminf convergence even if $\eta_0 = 0$ — as long as in the latter case we are careful when we increase Δ_k.
- If $\rho_k < \eta_1$ only reduce Δ_k if FQ.
- Criticality test includes curvature information
- Relatively general framework — which we can satisfy in a finite number of steps
- Do not require FQ/FL at every step
- Four possible outcomes: Successful, Acceptable, Model Improving, and Unsuccesful.
- After criticality step $\Delta_k = \min\{\tilde{\Delta}_k, \Delta_k\}$: good efficiency.
- All details of the proof are given

Convergence Highlights

- Convergence and not liminf convergence even if $\eta_0 = 0$ — as long as in the latter case we are careful when we increase Δ_k.
- If $\rho_k < \eta_1$ only reduce Δ_k if FQ.
- Criticality test includes curvature information
- Relatively general framework — which we can satisfy in a finite number of steps
- Do not require FQ/FL at every step
- Four possible outcomes: Successful, Acceptable, Model Improving, and Unsuccesful.
- After criticality step $\Delta_k = \min\{\tilde{\Delta}_k, \Delta_k\}$: good efficiency.
- All details of the proof are given

Convergence Highlights

- Convergence and not liminf convergence even if $\eta_0 = 0$ — as long as in the latter case we are careful when we increase Δ_k.
- If $\rho_k < \eta_1$ only reduce Δ_k if FQ.
- Criticality test includes curvature information
- Relatively general framework — which we can satisfy in a finite number of steps
- Do not require FQ/FL at every step
- Four possible outcomes: Successful, Acceptable, Model Improving, and Unsuccesful.
- After criticality step $\Delta_k = \min\{\tilde{\Delta}_k, \Delta_k\}$: good efficiency.
- All details of the proof are given

2^{nd}-**order convergence** (Details)

Step A: Show that $\|\sigma_k^m - \sigma_k\| \leq \kappa_\sigma \Delta_k$

Uses Wielandt-Hoffman and FQ

$$\kappa_{eh}\,\Delta \geq \|\nabla^2 f(x) - \nabla^2 m(x)\| \geq \|\nabla^2 f(x) - \nabla^2 m(x)\|_F/\sqrt{n} \geq \left\{\sum_{i=1}^n |\lambda_i - \mu_i|^2/n\right\}^{\frac{1}{2}}$$

$$\geq \frac{|\lambda_n - \mu_n|}{\sqrt{n}} = \frac{|\lambda_{min}(\nabla^2 f(x)) - \lambda_{min}(\nabla^2 m(x))|}{\sqrt{n}}$$

2^{nd}-**order convergence** (Details)

Step A: Show that $\|\sigma_k^m - \sigma_k\| \leq \kappa_\sigma \Delta_k$

Uses Wielandt-Hoffman and FQ

$$\kappa_{eh}\,\Delta \geq \|\nabla^2 f(x) - \nabla^2 m(x)\| \geq \|\nabla^2 f(x) - \nabla^2 m(x)\|_F/\sqrt{n} \geq \left\{\sum_{i=1}^n |\lambda_i - \mu_i|^2/n\right\}^{\frac{1}{2}}$$

$$\geq \frac{|\lambda_n - \mu_n|}{\sqrt{n}} = \frac{|\lambda_{min}(\nabla^2 f(x)) - \lambda_{min}(\nabla^2 m(x))|}{\sqrt{n}}$$

2^{nd}-**order convergence** (Details)

Step A: Show that $\|\sigma_k^m - \sigma_k\| \leq \kappa_\sigma \Delta_k$

Uses Wielandt-Hoffman and FQ

$$\kappa_{eh}\,\Delta \geq \|\nabla^2 f(x) - \nabla^2 m(x)\| \geq \|\nabla^2 f(x) - \nabla^2 m(x)\|_F/\sqrt{n} \geq \left\{\sum_{i=1}^n |\lambda_i - \mu_i|^2/n\right\}^{\frac{1}{2}}$$

$$\geq \frac{|\lambda_n - \mu_n|}{\sqrt{n}} = \frac{|\lambda_{min}(\nabla^2 f(x)) - \lambda_{min}(\nabla^2 m(x))|}{\sqrt{n}}$$

2^{nd}-order convergence (Details)

Step A: Show that $\|\sigma_k^m - \sigma_k\| \leq \kappa_\sigma \Delta_k$

Uses Wielandt-Hoffman and FQ

$$\kappa_{eh}\,\Delta \geq \|\nabla^2 f(x) - \nabla^2 m(x)\| \geq \|\nabla^2 f(x) - \nabla^2 m(x)\|_F / \sqrt{n} \geq \left\{\sum_{i=1}^n |\lambda_i - \mu_i|^2 / n\right\}^{\frac{1}{2}}$$

$$\geq \frac{|\lambda_n - \mu_n|}{\sqrt{n}} = \frac{|\lambda_{min}(\nabla^2 f(x)) - \lambda_{min}(\nabla^2 m(x))|}{\sqrt{n}}$$

Convergence (continued)

Step B: Show that unless $\sigma_k = 0$ criticality satisfied in finite number of geometry improvements

Proof by contraction. Scale by α, check σ_k^m small enough for new model, if it keeps failing Δ_k and $\sigma_k^m \to 0$, then use previous result

Step C: If $\sigma_k^m > \kappa_2 \; \forall k, \; \Delta_k > \kappa_3$,

We first show that if m_k is fully quadratic (FQ) on $B(x_k; \Delta_k)$ and

$$\Delta_k \;\le\; \min\left[\frac{\sigma_k^m}{2\kappa_{bhm}}, \frac{\kappa\sigma_k^m(1-\eta_1)}{8\kappa_{ef}\Delta_{\max}}, \frac{\kappa\sigma_k^m(1-\eta_1)}{8\kappa_{ef}}\right],$$

then the k-th iteration is successful. Uses approx. CP/EP and model FQ

Now whenever Δ_k falls below a certain value the k-th iteration has to be either model improving or successful, and hence $\Delta_{k+1} \ge \Delta_k$. This implies that $\Delta_j \ge \Delta_k$ for all j

Step D: If only a finite number of successful iterations, $\lim_{i \to +\infty} x_k = x^*$

If $\sigma_k \ne 0$ finite number model improvement iterations before FQ $\Rightarrow \infty$ iterations acceptable/unsuccesful and $\Delta_{k+1} < \Delta_k \Rightarrow \Delta_k \to 0$. Now use $\sigma(x_j) \;\le\; \|\sigma(x_j) - \sigma(x_{i_j})\| + \|\sigma(x_{i_j}) - \sigma_{i_j}^m\| + \sigma_{i_j}^m$.

Convergence (continued)

Step B: Show that unless $\sigma_k = 0$ criticality satisfied in finite number of geometry improvements

Proof by contradiction. Scale by α, check σ_k^m small enough for new model, if it keeps failing Δ_k and $\sigma_k^m \longrightarrow 0$, then use previous result

Step C: If $\sigma_k^m > \kappa_2 \; \forall k, \; \Delta_k > \kappa_3$,

We first show that if m_k is fully quadratic (FQ) on $B(x_k; \Delta_k)$ and

$$\Delta_k \;\leq\; \min\left[\frac{\sigma_k^m}{2\kappa_{bhm}}, \frac{\kappa\sigma_k^m(1-\eta_1)}{8\kappa_{ef}\Delta_{\max}}, \frac{\kappa\sigma_k^m(1-\eta_1)}{8\kappa_{ef}}\right],$$

then the k-th iteration is successful. Uses approx. CP/EP and model FQ

Now whenever Δ_k falls below a certain value the k-th iteration has to be either model improving or successful, and hence $\Delta_{k+1} \geq \Delta_k$. This implies that $\Delta_j \geq \Delta_k$ for all j

Step D: If only a finite number of successful iterations, $\lim_{i \longrightarrow +\infty} x_k = x^*$

If $\sigma_k \neq 0$ finite number model improvement iterations before FQ $\Rightarrow \infty$ iterations acceptable/unsuccesful and $\Delta_{k+1} < \Delta_k \Rightarrow \Delta_k \longrightarrow 0$. Now use $\sigma(x_j) \;\leq\; \|\sigma(x_j) - \sigma(x_{i_j})\| + \|\sigma(x_{i_j}) - \sigma_{i_j}^m\| + \sigma_{i_j}^m$.

Convergence (continued)

Step B: Show that unless $\sigma_k = 0$ criticality satisfied in finite number of geometry improvements

Proof by contradiction. Scale by α, check σ_k^m small enough for new model, if it keeps failing Δ_k and $\sigma_k^m \to 0$, then use previous result

Step C: If $\sigma_k^m > \kappa_2 \; \forall k, \;\; \Delta_k > \kappa_3$,

We first show that if m_k is fully quadratic (FQ) on $B(x_k; \Delta_k)$ and

$$\Delta_k \;\leq\; \min\left[\frac{\sigma_k^m}{2\kappa_{bhm}}, \frac{\kappa\sigma_k^m(1-\eta_1)}{8\kappa_{ef}\Delta_{\max}}, \frac{\kappa\sigma_k^m(1-\eta_1)}{8\kappa_{ef}}\right],$$

then the k-th iteration is successful. Uses approx. CP/EP and model FQ

Now whenever Δ_k falls below a certain value the k-th iteration has to be either model improving or successful, and hence $\Delta_{k+1} \geq \Delta_k$. This implies that $\Delta_j \geq \Delta_k$ for all j

Step D: If only a finite number of successful iterations, $\lim_{i \to +\infty} x_k = x^*$

If $\sigma_k \neq 0$ finite number model improvement iterations before FQ $\Rightarrow$ ∞ iterations acceptable/unsuccesful and $\Delta_{k+1} < \Delta_k \Rightarrow \Delta_k \to 0$. Now use $\sigma(x_j) \;\leq\; \|\sigma(x_j) - \sigma(x_{i_j})\| + \|\sigma(x_{i_j}) - \sigma_{i_j}^m\| + \sigma_{i_j}^m$.

Convergence (continued)

Step B: Show that unless $\sigma_k = 0$ criticality satisfied in finite number of geometry improvements

Proof by contradiction. Scale by α, check σ_k^m small enough for new model, if it keeps failing Δ_k and $\sigma_k^m \to 0$, then use previous result

Step C: If $\sigma_k^m > \kappa_2 \; \forall k, \;\; \Delta_k > \kappa_3$,

We first show that if m_k is fully quadratic (FQ) on $B(x_k; \Delta_k)$ and

$$\Delta_k \;\leq\; \min\left[\frac{\sigma_k^m}{2\kappa_{bhm}}, \frac{\kappa\sigma_k^m(1-\eta_1)}{8\kappa_{ef}\Delta_{\max}}, \frac{\kappa\sigma_k^m(1-\eta_1)}{8\kappa_{ef}}\right],$$

then the k-th iteration is successful. Uses approx. CP/EP and model FQ

Now whenever Δ_k falls below a certain value the k-th iteration has to be either model improving or successful, and hence $\Delta_{k+1} \geq \Delta_k$. This implies that $\Delta_j \geq \Delta_k$ for all j

Step D: If only a finite number of successful iterations, $\lim_{i \to +\infty} x_k = x^*$

If $\sigma_k \neq 0$ finite number model improvement iterations before FQ $\Rightarrow$ ∞ iterations acceptable/unsuccesful and $\Delta_{k+1} < \Delta_k \Rightarrow \Delta_k \to 0$. Now use $\sigma(x_j) \;\leq\; \|\sigma(x_j) - \sigma(x_{i_j})\| + \|\sigma(x_{i_j}) - \sigma_{i_j}^m\| + \sigma_{i_j}^m$.

Convergence (continued)

Step B: Show that unless $\sigma_k = 0$ criticality satisfied in finite number of geometry improvements

Proof by contradiction. Scale by α, check σ_k^m small enough for new model, if it keeps failing Δ_k and $\sigma_k^m \to 0$, then use previous result

Step C: If $\sigma_k^m > \kappa_2 \; \forall k, \;\; \Delta_k > \kappa_3$,

We first show that if m_k is fully quadratic (FQ) on $B(x_k; \Delta_k)$ and

$$\Delta_k \;\leq\; \min\left[\frac{\sigma_k^m}{2\kappa_{bhm}}, \frac{\kappa\sigma_k^m(1-\eta_1)}{8\kappa_{ef}\Delta_{\max}}, \frac{\kappa\sigma_k^m(1-\eta_1)}{8\kappa_{ef}}\right],$$

then the k-th iteration is successful. Uses approx. CP/EP and model FQ

Now whenever Δ_k falls below a certain value the k-th iteration has to be either model improving or successful, and hence $\Delta_{k+1} \geq \Delta_k$. This implies that $\Delta_j \geq \Delta_k$ for all j

Step D: If only a finite number of successful iterations, $\lim_{i \to +\infty} x_k = x^*$

If $\sigma_k \neq 0$ finite number model improvement iterations before FQ $\Rightarrow \infty$ iterations acceptable/unsuccesful and $\Delta_{k+1} < \Delta_k \Rightarrow \Delta_k \to 0$. Now use $\sigma(x_j) \;\leq\; \|\sigma(x_j) - \sigma(x_{i_j})\| + \|\sigma(x_{i_j}) - \sigma_{i_j}^m\| + \sigma_{i_j}^m$.

Convergence (continued)

Step B: Show that unless $\sigma_k = 0$ criticality satisfied in finite number of geometry improvements

Proof by contradiction. Scale by α, check σ_k^m small enough for new model, if it keeps failing Δ_k and $\sigma_k^m \to 0$, then use previous result

Step C: If $\sigma_k^m > \kappa_2 \ \forall k, \ \ \Delta_k > \kappa_3$,

We first show that if m_k is fully quadratic (FQ) on $B(x_k; \Delta_k)$ and

$$\Delta_k \ \leq \ \min \left[\frac{\sigma_k^m}{2\kappa_{bhm}}, \frac{\kappa \sigma_k^m (1-\eta_1)}{8\kappa_{ef}\Delta_{\max}}, \frac{\kappa \sigma_k^m (1-\eta_1)}{8\kappa_{ef}} \right],$$

then the k-th iteration is successful. Uses approx. CP/EP and model FQ

Now whenever Δ_k falls below a certain value the k-th iteration has to be either model improving or successful, and hence $\Delta_{k+1} \geq \Delta_k$. This implies that $\Delta_j \geq \Delta_k$ for all j

Step D: If only a finite number of successful iterations, $\lim_{i \to +\infty} x_k = x^*$

If $\sigma_k \neq 0$ finite number model improvement iterations before FQ $\Rightarrow \infty$ iterations acceptable/unsuccesful and $\Delta_{k+1} < \Delta_k \Rightarrow \Delta_k \to 0$. Now use $\sigma(x_j) \ \leq \ \|\sigma(x_j) - \sigma(x_{i_j})\| + \|\sigma(x_{i_j}) - \sigma_{i_j}^m\| + \sigma_{i_j}^m$.

Convergence (continued)

Step B: Show that unless $\sigma_k = 0$ criticality satisfied in finite number of geometry improvements

Proof by contradiction. Scale by α, check σ_k^m small enough for new model, if it keeps failing Δ_k and $\sigma_k^m \to 0$, then use previous result

Step C: If $\sigma_k^m > \kappa_2 \; \forall k, \;\; \Delta_k > \kappa_3$,

We first show that if m_k is fully quadratic (FQ) on $B(x_k; \Delta_k)$ and

$$\Delta_k \;\leq\; \min\left[\frac{\sigma_k^m}{2\kappa_{bhm}}, \frac{\kappa\sigma_k^m(1-\eta_1)}{8\kappa_{ef}\Delta_{\max}}, \frac{\kappa\sigma_k^m(1-\eta_1)}{8\kappa_{ef}}\right],$$

then the k-th iteration is successful. Uses approx. CP/EP and model FQ

Now whenever Δ_k falls below a certain value the k-th iteration has to be either model improving or successful, and hence $\Delta_{k+1} \geq \Delta_k$. This implies that $\Delta_j \geq \Delta_k$ for all j

Step D: If only a finite number of successful iterations, $\lim_{i \to +\infty} x_k = x^*$

If $\sigma_k \neq 0$ finite number model improvement iterations before FQ $\Rightarrow \infty$ iterations acceptable/unsuccesful and $\Delta_{k+1} < \Delta_k \Rightarrow \Delta_k \to 0$. Now use
$\sigma(x_j) \;\leq\; \|\sigma(x_j) - \sigma(x_{i_j})\| + \|\sigma(x_{i_j}) - \sigma_{i_j}^m\| + \sigma_{i_j}^m$.

Convergence (continued)

Step E: Infinite number of successful iterations ensures $\liminf_{k\to\infty} \sigma_k^m = 0$

By contradiction. If for all k,

$$\sigma_k^m \;\geq\; \kappa_1$$

then by Step C for each successful iteration

$$\begin{aligned} f(x_k) - f(x_{k+1}) \;&\geq\; \eta_1(m(x_k) - m(x_k + s_k)) \;\geq \\ \eta_1 \frac{\kappa_{fod}}{2} \max\left\{ \|g_k\| \min\left[\frac{\|g_k\|}{\kappa_{bhm}}, \Delta_k\right], -\tau_k \Delta_k^2 \right\}. \end{aligned}$$

Since $\sigma_k^m \geq \kappa_1 \Rightarrow$ rhs is bounded away from zero for all k, and, hence, so is $f(x_{k+1}) - f(x_k)$ for each successful iteration. The number of successful iterates cannot be infinite since f is bounded from below.

Step F: Existence of subsequence $\lim_{i\to\infty} \sigma_{k_i} = 0$ provided $\lim_{i\to\infty} \sigma_{k_i}^m = 0$

Uses $\sigma_{k_i} \leq \sigma_{k_i}^m + |\sigma_{k_i}^m - \sigma_{k_i}|$, the Step A result and CSA

Steps E and F together gives global convergence to a subsequence of successful iterates

Convergence (continued)

Step E: Infinite number of successful iterations ensures $\liminf_{k\to\infty} \sigma_k^m = 0$

By contradiction. If for all k,

$$\sigma_k^m \ \geq \ \kappa_1$$

then by Step C for each successful iteration

$$f(x_k) - f(x_{k+1}) \ \geq \ \eta_1(m(x_k) - m(x_k + s_k)) \ \geq$$
$$\eta_1 \frac{\kappa_{fod}}{2} \max\left\{ \|g_k\| \min\left[\frac{\|g_k\|}{\kappa_{bhm}}, \Delta_k\right], -\tau_k \Delta_k^2 \right\}.$$

Since $\sigma_k^m \geq \kappa_1 \Rightarrow$ rhs is bounded away from zero for all k, and, hence, so is $f(x_{k+1}) - f(x_k)$ for each successful iteration. The number of successful iterates cannot be infinite since f is bounded from below.

Step F: Existence of subsequence $\lim_{i\to\infty} \sigma_{k_i} = 0$ provided $\lim_{i\to\infty} \sigma_{k_i}^m = 0$

Uses $\sigma_{k_i} \leq \sigma_{k_i}^m + |\sigma_{k_i}^m - \sigma_{k_i}|$, the Step A result and CSA

Steps E and F together gives global convergence to a subsequence of successful iterates

Convergence (continued)

Step E: Infinite number of successful iterations ensures $\liminf_{k\to\infty} \sigma_k^m = 0$

By contradiction. If for all k,

$$\sigma_k^m \;\geq\; \kappa_1$$

then by Step C for each successful iteration

$$f(x_k) - f(x_{k+1}) \;\geq\; \eta_1 (m(x_k) - m(x_k + s_k)) \;\geq\; \eta_1 \frac{\kappa_{fod}}{2} \max\left\{ \|g_k\| \min\left[\frac{\|g_k\|}{\kappa_{bhm}}, \Delta_k \right], -\tau_k \Delta_k^2 \right\}.$$

Since $\sigma_k^m \geq \kappa_1 \Rightarrow$ rhs is bounded away from zero for all k, and, hence, so is $f(x_{k+1}) - f(x_k)$ for each successful iteration. The number of successful iterates cannot be infinite since f is bounded from below.

Step F: Existence of subsequence $\lim_{i\to\infty} \sigma_{k_i} = 0$ provided $\lim_{i\to\infty} \sigma_{k_i}^m = 0$

Uses $\sigma_{k_i} \leq \sigma_{k_i}^m + |\sigma_{k_i}^m - \sigma_{k_i}|$, the Step A result and CSA

Steps E and F together gives global convergence to a subsequence of successful iterates

Convergence (continued)

Step E: Infinite number of successful iterations ensures $\liminf_{k\to\infty} \sigma_k^m = 0$

By contradiction. If for all k,

$$\sigma_k^m \;\geq\; \kappa_1$$

then by Step C for each successful iteration

$$f(x_k) - f(x_{k+1}) \;\geq\; \eta_1(m(x_k) - m(x_k + s_k)) \;\geq\; \eta_1 \frac{\kappa_{fod}}{2} \max\left\{ \|g_k\| \min\left[\frac{\|g_k\|}{\kappa_{bhm}}, \Delta_k \right], -\tau_k \Delta_k^2 \right\}.$$

Since $\sigma_k^m \geq \kappa_1 \Rightarrow$ rhs is bounded away from zero for all k, and, hence, so is $f(x_{k+1}) - f(x_k)$ for each successful iteration. The number of successful iterates cannot be infinite since f is bounded from below.

Step F: Existence of subsequence $\lim_{i\to\infty} \sigma_{k_i} = 0$ provided $\lim_{i\to\infty} \sigma_{k_i}^m = 0$

Uses $\sigma_{k_i} \leq \sigma_{k_i}^m + |\sigma_{k_i}^m - \sigma_{k_i}|$, the Step A result and CSA

Steps E and F together gives global convergence to a subsequence of successful iterates

Convergence (continued)

Step E: Infinite number of successful iterations ensures $\liminf_{k\to\infty} \sigma_k^m = 0$

By contradiction. If for all k,

$$\sigma_k^m \geq \kappa_1$$

then by Step C for each successful iteration

$$f(x_k) - f(x_{k+1}) \geq \eta_1 (m(x_k) - m(x_k + s_k)) \geq \eta_1 \frac{\kappa_{fod}}{2} \max\left\{ \|g_k\| \min\left[\frac{\|g_k\|}{\kappa_{bhm}}, \Delta_k \right], -\tau_k \Delta_k^2 \right\}.$$

Since $\sigma_k^m \geq \kappa_1 \Rightarrow$ rhs is bounded away from zero for all k, and, hence, so is $f(x_{k+1}) - f(x_k)$ for each successful iteration. The number of successful iterates cannot be infinite since f is bounded from below.

Step F: Existence of subsequence $\lim_{i\to\infty} \sigma_{k_i} = 0$ provided $\lim_{i\to\infty} \sigma_{k_i}^m = 0$

Uses $\sigma_{k_i} \leq \sigma_{k_i}^m + |\sigma_{k_i}^m - \sigma_{k_i}|$, the Step A result and CSA

Steps E and F together gives global convergence to a subsequence of successful iterates

Convergence (continued)

Step G: Assuming either $\eta_0 = \eta_1 \in (0,1)$ or Δ_k is only increased if $\Delta_k < \beta\sigma_k^m$. Then, $\lim_{k\to+\infty} \sigma(x_k) = 0$.

By contradiction. Suppose there exists a subsequence $\{k_i\}$ of successful or acceptable iterations $\sigma_{k_i} \geq \epsilon_0 > 0$. (Ignore the other iterations, since for them x_k does not change). Then, Step F $\Rightarrow \sigma_{k_i}^m \geq 2\epsilon > 0$. Pick ϵ such that $(2+\kappa_\sigma\mu)\epsilon \leq \frac{1}{2}\epsilon_0$. Step E then ensures the existence, for each k_i in the subsequence, of a first iteration $\ell_i > k_i$ such that m_{ℓ_i} is FQ and $\sigma_{\ell_i} < \epsilon \Rightarrow$ another subsequence indexed by $\{\ell_i\}$ such that $\sigma_k \geq \epsilon$ for $k_i \leq k < \ell_i$ and $\sigma_{\ell_i} < \epsilon$, for sufficiently large i. Restricting attention to the subsequence of successful or acceptable iterations whose indices are in the set $\mathcal{K} = \{k \in \mathbb{N}_0 : k_i \leq k < \ell_i\}$, we show that (using the bound on the CP/EP decrease and the fact that f is bounded below) $\lim_{\substack{k\to+\infty \\ k\in\mathcal{K}}} \Delta_k = 0$ and for sufficiently large k they all are successful ones such that $\Delta_k < \beta\sigma_k^m$ and so

$$\|x_{k_i} - x_{\ell_i}\| \leq \sum_{j=k_i}^{\ell_i-1} \Delta_j \leq \sum_{j=k_i}^{\ell_i-1} \gamma_+^{\ell_i-j}\Delta_j \leq \frac{\gamma_+}{\gamma_+ - 1}\Delta_{\ell_i-1}.$$

. Finally use

$$\|\sigma(x_{k_i})\| \leq \|\sigma(x_{k_i}) - \sigma(x_{\ell_i})\| + \|\sigma(x_{\ell_i}) - \sigma_{\ell_i}^m\| + \|\sigma_{\ell_i}^m\|.$$

Convergence (continued)

Step G: Assuming either $\eta_0 = \eta_1 \in (0,1)$ or Δ_k is only increased if $\Delta_k < \beta \sigma_k^m$. Then, $\lim_{k \to +\infty} \sigma(x_k) = 0$.

By contradiction. Suppose there exists a subsequence $\{k_i\}$ of successful or acceptable iterations $\sigma_{k_i} \geq \epsilon_0 > 0$. (Ignore the other iterations, since for them x_k does not change). Then, Step F $\Rightarrow \sigma_{k_i}^m \geq 2\epsilon > 0$. Pick ϵ such that $(2 + \kappa_\sigma \mu)\epsilon \leq \frac{1}{2}\epsilon_0$. Step E then ensures the existence, for each k_i in the subsequence, of a first iteration $\ell_i > k_i$ such that m_{ℓ_i} is FQ and $\sigma_{\ell_i} < \epsilon \Rightarrow$ another subsequence indexed by $\{\ell_i\}$ such that $\sigma_k \geq \epsilon$ for $k_i \leq k < \ell_i$ and $\sigma_{\ell_i} < \epsilon$, for sufficiently large i. Restricting attention to the subsequence of successful or acceptable iterations whose indices are in the set $\mathcal{K} = \{k \in \mathbb{N}_0 : k_i \leq k < \ell_i\}$, we show that (using the bound on the CP/EP decrease and the fact that f is bounded below) $\lim_{\substack{k \to +\infty \\ k \in \mathcal{K}}} \Delta_k = 0$ and for sufficiently large k they all are successful ones such that $\Delta_k < \beta \sigma_k^m$ and so

$$\|x_{k_i} - x_{\ell_i}\| \leq \sum_{j=k_i}^{\ell_i - 1} \Delta_j \leq \sum_{j=k_i}^{\ell_i - 1} \gamma_+^{\ell_i - j} \Delta_j \leq \frac{\gamma_+}{\gamma_+ - 1} \Delta_{\ell_i - 1}.$$

. Finally use

$$\|\sigma(x_{k_i})\| \leq \|\sigma(x_{k_i}) - \sigma(x_{\ell_i})\| + \|\sigma(x_{\ell_i}) - \sigma_{\ell_i}^m\| + \|\sigma_{\ell_i}^m\|.$$

Viability in Practise

Linear case

Let f be a function defined from $\mathbb{R}^n$ to $\mathbb{R}$.
Consider $p = n + 1$ interpolation points:

$$Y \;=\; \{y^0, \ldots, y^{p-1}\} \;\subset\; B(\Delta)$$

in a closed ball $B(\Delta)$ of radius $\Delta > 0$.
Given these $p = n + 1$ points we can aim to built the fully linear interpolation model:

$$m(x) \;=\; c + g^\top x \;=\; c + \sum_{1 \le k \le n} g_k x_k .$$

Viability in Practise

Linear case

Let f be a function defined from $\mathbb{R}^n$ to $\mathbb{R}$.
Consider $p = n + 1$ interpolation points:

$$Y \;=\; \{y^0, \ldots, y^{p-1}\} \;\subset\; B(\Delta)$$

in a closed ball $B(\Delta)$ of radius $\Delta > 0$.
Given these $p = n + 1$ points we can aim to built the fully linear interpolation model:

$$m(x) \;=\; c + g^\top x \;=\; c + \sum_{1 \le k \le n} g_k x_k .$$

Viability in Practise

Linear case

Let f be a function defined from $\mathbb{R}^n$ to $\mathbb{R}$.
Consider $p = n + 1$ interpolation points:

$$Y \;=\; \{y^0, \ldots, y^{p-1}\} \;\subset\; B(\Delta)$$

in a closed ball $B(\Delta)$ of radius $\Delta > 0$.
Given these $p = n + 1$ points we can aim to built the fully linear interpolation model:

$$m(x) \;=\; c + g^\top x \;=\; c + \sum_{1 \le k \le n} g_k x_k .$$

Viability in Practise

Linear case

Let f be a function defined from $\mathbb{R}^n$ to $\mathbb{R}$.
Consider $p = n + 1$ interpolation points:

$$Y \;=\; \left\{y^0, \ldots, y^{p-1}\right\} \;\subset\; B(\Delta)$$

in a closed ball $B(\Delta)$ of radius $\Delta > 0$.
Given these $p = n + 1$ points we can aim to built the fully linear interpolation model:

$$m(x) \;=\; c + g^\top x \;=\; c + \sum_{1 \le k \le n} g_k x_k .$$

Linear case (continued)

The coefficients c, $g_1, \ldots, g_n$ are defined by the interpolating conditions:

$$m(y^i) \;=\; f(y^i), \; i = 0, \ldots, n.$$

If the coefficient matrix

$$\begin{bmatrix} 1 & y_1^0 & \cdots & y_n^0 \\ \vdots & \vdots & & \vdots \\ 1 & y_1^n & \cdots & y_n^n \end{bmatrix}$$

of the system is nonsingular the set of points Y is said to poised, and the polynomial coefficients are well defined.
Otherwise, we say that the set Y is non-poised.
From a numerical point of view we care how close we are to non-poised!.

Linear case (continued)

The coefficients c, $g_1, \ldots, g_n$ are defined by the interpolating conditions:

$$m(y^i) \;=\; f(y^i), \; i = 0, \ldots, n.$$

If the coefficient matrix

$$\begin{bmatrix} 1 & y_1^0 & \cdots & y_n^0 \\ \vdots & \vdots & & \vdots \\ 1 & y_1^n & \cdots & y_n^n \end{bmatrix}$$

of the system is nonsingular the set of points Y is said to poised, and the polynomial coefficients are well defined.

Otherwise, we say that the set Y is non-poised.

From a numerical point of view we care how close we are to non-poised!.

Linear case (continued)

The coefficients c, $g_1, \ldots, g_n$ are defined by the interpolating conditions:

$$m(y^i) \; = \; f(y^i), \; i = 0, \ldots, n.$$

If the coefficient matrix

$$\begin{bmatrix} 1 & y_1^0 & \cdots & y_n^0 \\ \vdots & \vdots & & \vdots \\ 1 & y_1^n & \cdots & y_n^n \end{bmatrix}$$

of the system is nonsingular the set of points Y is said to poised, and the polynomial coefficients are well defined.
Otherwise, we say that the set Y is non-poised.
From a numerical point of view we care how close we are to non-poised!.

Linear case (continued)

The coefficients c, $g_1, \ldots, g_n$ are defined by the interpolating conditions:

$$m(y^i) \; = \; f(y^i), \; i = 0, \ldots, n.$$

If the coefficient matrix

$$\begin{bmatrix} 1 & y_1^0 & \cdots & y_n^0 \\ \vdots & \vdots & & \vdots \\ 1 & y_1^n & \cdots & y_n^n \end{bmatrix}$$

of the system is nonsingular the set of points Y is said to poised, and the polynomial coefficients are well defined.
Otherwise, we say that the set Y is non-poised.
From a numerical point of view we care how close we are to non-poised!.

Linear case (continued)

Consider a point x in the ball $B(\Delta)$ and the errors:

$$m(x) \;=\; f(x) + e^f(x), \qquad g \;=\; \nabla f(x) + e^g(x).$$

Subtracting $m(x) = f(x) + e^f(x)$ from all $m(y^i) = f(y^i),\ i = 0, \ldots, n$:

$$(y^i - x)^\top g \;=\; f(y^i) - f(x) - e^f(x), \quad i = 0, \ldots, n.$$

Expanding f by Taylor:

$$(y^i - x)^\top e^g(x) \;=\; \mathcal{O}(\Delta^2) - e^f(x), \quad i = 0, \ldots, n.$$

Subtracting these equations in pairs,

$$(y^i - y^j)^\top e^g(x) \;=\; \mathcal{O}(\Delta^2), \quad 0 \le i < j \le n,$$

Linear case (continued)

Consider a point x in the ball $B(\Delta)$ and the errors:

$$m(x) \;=\; f(x) + e^f(x), \qquad g \;=\; \nabla f(x) + e^g(x).$$

Subtracting $m(x) = f(x) + e^f(x)$ from all $m(y^i) = f(y^i),\ i = 0, \ldots, n$:

$$(y^i - x)^\top g \;=\; f(y^i) - f(x) - e^f(x), \qquad i = 0, \ldots, n.$$

Expanding f by Taylor:

$$(y^i - x)^\top e^g(x) \;=\; \mathcal{O}(\Delta^2) - e^f(x), \qquad i = 0, \ldots, n.$$

Subtracting these equations in pairs,

$$(y^i - y^j)^\top e^g(x) \;=\; \mathcal{O}(\Delta^2), \qquad 0 \le i < j \le n,$$

Linear case (continued)

Consider a point x in the ball $B(\Delta)$ and the errors:

$$m(x) \;=\; f(x) + e^f(x), \qquad g \;=\; \nabla f(x) + e^g(x).$$

Subtracting $m(x) = f(x) + e^f(x)$ from all
$m(y^i) = f(y^i),\ i = 0, \ldots, n$:

$$(y^i - x)^\top g \;=\; f(y^i) - f(x) - e^f(x), \quad i = 0, \ldots, n.$$

Expanding f by Taylor:

$$(y^i - x)^\top e^g(x) \;=\; \mathcal{O}(\Delta^2) - e^f(x), \quad i = 0, \ldots, n.$$

Subtracting these equations in pairs,

$$(y^i - y^j)^\top e^g(x) \;=\; \mathcal{O}(\Delta^2), \quad 0 \le i < j \le n,$$

Linear case (continued)

Consider a point x in the ball $B(\Delta)$ and the errors:

$$m(x) \;=\; f(x) + e^f(x), \qquad g \;=\; \nabla f(x) + e^g(x).$$

Subtracting $m(x) = f(x) + e^f(x)$ from all
$m(y^i) = f(y^i),\; i = 0, \ldots, n$:

$$(y^i - x)^\top g \;=\; f(y^i) - f(x) - e^f(x), \quad i = 0, \ldots, n.$$

Expanding f by Taylor:

$$(y^i - x)^\top e^g(x) \;=\; \mathcal{O}(\Delta^2) - e^f(x), \quad i = 0, \ldots, n.$$

Subtracting these equations in pairs,

$$(y^i - y^j)^\top e^g(x) \;=\; \mathcal{O}(\Delta^2), \quad 0 \le i < j \le n,$$

Error estimates in the quadratic case

THEOREM: Given $Y = \{0, y^1, \ldots, y^p\} \subset B(\Delta)$, $p = (n+1)(n+2)/2 - 1$, and f be $C^2(\gamma)$ in $\Omega \supset B(\Delta)$. Let $\hat{Q}_{p+1 \times p+1}$ be the scaled matrix.

Then, for all points x in $B(\Delta)$, we have that

$$\|\nabla^2 f(x) - \nabla^2 m(x)\|_2 \;\leq\; \left(c^h p^{\frac{1}{2}} \gamma \|\hat{Q}^{-1}_{p \times p}\| \right) \Delta,$$

$$\|\nabla f(x) - \nabla m(x)\|_2 \;\leq\; \left(c^g p^{\frac{1}{2}} \gamma \|\hat{Q}^{-1}_{p \times p}\| \right) \Delta^2,$$

and

$$|f(x) - m(x)| \;\leq\; \left(c^f p^{\frac{1}{2}} \gamma \|\hat{Q}^{-1}_{p \times p}\| + \gamma/6 \right) \Delta^3.$$

The constants are $c^h = 3\sqrt{2}/2$, $c^g = 3(1+\sqrt{2})/2$, and $c^f = (6 + 9\sqrt{2})/4$.

Error estimates in the quadratic case

THEOREM: Given $Y = \{0, y^1, \ldots, y^p\} \subset B(\Delta)$, $p = (n+1)(n+2)/2 - 1$, and f be $C^2(\gamma)$ in $\Omega \supset B(\Delta)$. Let $\hat{Q}_{p+1 \times p+1}$ be the scaled matrix.

Then, for all points x in $B(\Delta)$, we have that

$$\|\nabla^2 f(x) - \nabla^2 m(x)\|_2 \;\leq\; \left(c^h p^{\frac{1}{2}} \gamma \|\hat{Q}^{-1}_{p \times p}\| \right) \Delta,$$

$$\|\nabla f(x) - \nabla m(x)\|_2 \;\leq\; \left(c^g p^{\frac{1}{2}} \gamma \|\hat{Q}^{-1}_{p \times p}\| \right) \Delta^2,$$

and

$$|f(x) - m(x)| \;\leq\; \left(c^f p^{\frac{1}{2}} \gamma \|\hat{Q}^{-1}_{p \times p}\| + \gamma/6 \right) \Delta^3.$$

The constants are $c^h = 3\sqrt{2}/2$, $c^g = 3(1+\sqrt{2})/2$, and $c^f = (6 + 9\sqrt{2})/4$.

Error estimates in the quadratic case

THEOREM: Given $Y = \{0, y^1, \ldots, y^p\} \subset B(\Delta)$, $p = (n+1)(n+2)/2 - 1$, and f be $C^2(\gamma)$ in $\Omega \supset B(\Delta)$. Let $\hat{Q}_{p+1 \times p+1}$ be the scaled matrix.

Then, for all points x in $B(\Delta)$, we have that

$$\|\nabla^2 f(x) - \nabla^2 m(x)\|_2 \;\leq\; \left(c^h p^{\frac{1}{2}} \gamma \|\hat{Q}^{-1}_{p \times p}\| \right) \Delta,$$

$$\|\nabla f(x) - \nabla m(x)\|_2 \;\leq\; \left(c^g p^{\frac{1}{2}} \gamma \|\hat{Q}^{-1}_{p \times p}\| \right) \Delta^2,$$

and

$$|f(x) - m(x)| \;\leq\; \left(c^f p^{\frac{1}{2}} \gamma \|\hat{Q}^{-1}_{p \times p}\| + \gamma/6 \right) \Delta^3.$$

The constants are $c^h = 3\sqrt{2}/2$, $c^g = 3(1+\sqrt{2})/2$, and $c^f = (6+9\sqrt{2})/4$.

Error estimates in the quadratic case

THEOREM: Given $Y = \{0, y^1, \ldots, y^p\} \subset B(\Delta)$, $p = (n+1)(n+2)/2 - 1$, and f be $C^2(\gamma)$ in $\Omega \supset B(\Delta)$. Let $\hat{Q}_{p+1\times p+1}$ be the scaled matrix.

Then, for all points x in $B(\Delta)$, we have that

$$\|\nabla^2 f(x) - \nabla^2 m(x)\|_2 \;\leq\; \left(c^h p^{\frac{1}{2}} \gamma \| \hat{Q}^{-1}_{p\times p} \| \right) \Delta,$$

$$\|\nabla f(x) - \nabla m(x)\|_2 \;\leq\; \left(c^g p^{\frac{1}{2}} \gamma \| \hat{Q}^{-1}_{p\times p} \| \right) \Delta^2,$$

and

$$|f(x) - m(x)| \;\leq\; \left(c^f p^{\frac{1}{2}} \gamma \| \hat{Q}^{-1}_{p\times p} \| + \gamma/6 \right) \Delta^3.$$

The constants are $c^h = 3\sqrt{2}/2$, $c^g = 3(1+\sqrt{2})/2$, and $c^f = (6+9\sqrt{2})/4$.

Error estimates in the quadratic case

THEOREM: Given $Y = \{0, y^1, \ldots, y^p\} \subset B(\Delta)$, $p = (n+1)(n+2)/2 - 1$, and f be $C^2(\gamma)$ in $\Omega \supset B(\Delta)$. Let $\hat{Q}_{p+1\times p+1}$ be the scaled matrix.

Then, for all points x in $B(\Delta)$, we have that

$$\|\nabla^2 f(x) - \nabla^2 m(x)\|_2 \;\leq\; \left(c^h p^{\frac{1}{2}} \gamma \|\hat{Q}_{p\times p}^{-1}\|\right) \Delta,$$

$$\|\nabla f(x) - \nabla m(x)\|_2 \;\leq\; \left(c^g p^{\frac{1}{2}} \gamma \|\hat{Q}_{p\times p}^{-1}\|\right) \Delta^2,$$

and

$$|f(x) - m(x)| \;\leq\; \left(c^f p^{\frac{1}{2}} \gamma \|\hat{Q}_{p\times p}^{-1}\| + \gamma/6\right) \Delta^3.$$

The constants are $c^h = 3\sqrt{2}/2$, $c^g = 3(1+\sqrt{2})/2$, and $c^f = (6 + 9\sqrt{2})/4$.

Ensuring well-poisedness (quadratic case)

Consider the case $n = 2$ (and $p = 6$):

$$
\begin{aligned}
Q_{5\times 5} \quad &= \quad \begin{bmatrix}
y_1^1 - y_1^0 & y_2^1 - y_2^0 & \frac{1}{2}(y_1^1)^2 - \frac{1}{2}(y_1^0)^2 & \frac{1}{2}(y_2^1)^2 - \frac{1}{2}(y_2^1)^2 & y_1^1 y_2^1 - y_1^0 y_2^0 \\
y_1^2 - y_1^0 & y_2^2 - y_2^0 & \frac{1}{2}(y_1^2)^2 - \frac{1}{2}(y_1^0)^2 & \frac{1}{2}(y_2^2)^2 - \frac{1}{2}(y_2^0)^2 & y_1^2 y_2^2 - y_1^0 y_2^0 \\
y_1^3 - y_1^0 & y_2^3 - y_2^0 & \frac{1}{2}(y_1^3)^2 - \frac{1}{2}(y_1^0)^2 & \frac{1}{2}(y_2^3)^2 - \frac{1}{2}(y_2^0)^2 & y_1^3 y_2^3 - y_1^0 y_2^0 \\
y_1^4 - y_1^0 & y_2^4 - y_2^0 & \frac{1}{2}(y_1^4)^2 - \frac{1}{2}(y_1^0)^2 & \frac{1}{2}(y_2^4)^2 - \frac{1}{2}(y_2^0)^2 & y_1^4 y_2^4 - y_1^0 y_2^0 \\
y_1^5 - y_1^0 & y_2^5 - y_2^0 & \frac{1}{2}(y_1^5)^2 - \frac{1}{2}(y_1^0)^2 & \frac{1}{2}(y_2^5)^2 - \frac{1}{2}(y_2^0)^2 & y_1^5 y_2^5 - y_1^0 y_2^0
\end{bmatrix} \\
&= \quad \begin{bmatrix}
\phi(y^1)^\top \\
\phi(y^2)^\top \\
\vdots \\
\phi(y^5)^\top
\end{bmatrix} \qquad \text{given that } y^0 = 0 \text{ without loss of generality.}
\end{aligned}
$$

Remember

$$
\phi(x) \; = \; \begin{bmatrix} x_1 & x_2 & \frac{x_1^2}{2} & \frac{x_2^2}{2} & x_1 x_2 \end{bmatrix}^\top .
$$

Ensuring well-poisedness (continued)

Consider a scaled version of $Q_{5\times 5}$:

$$\hat{Q}_{5\times 5} = Q_{5\times 5} \begin{bmatrix} \Delta^{-1} & 0 & 0 & 0 & 0 \\ 0 & \Delta^{-1} & 0 & 0 & 0 \\ 0 & 0 & \Delta^{-2} & 0 & 0 \\ 0 & 0 & 0 & \Delta^{-2} & 0 \\ 0 & 0 & 0 & 0 & \Delta^{-2} \end{bmatrix} = \begin{bmatrix} \phi(y^1/\Delta)^\top \\ \phi(y^2/\Delta)^\top \\ \vdots \\ \phi(y^5/\Delta)^\top \end{bmatrix}.$$

Our goal is compute $\hat{Q}_{5\times 5}$ such that its inverse has the smallest possible norm:

$$\min \ \|(\hat{Q}_{5\times 5})^{-1}\|_2 \quad \Longleftrightarrow \quad \min \ \Lambda.$$

IDEA: Factorize the matrix $\hat{Q}_{5\times 5}$ by rows, changing the points in Y, if necessary, so that the pivots become as large as possible.

Ensuring well-poisedness (continued)

Consider a scaled version of $Q_{5\times 5}$:

$$\hat{Q}_{5\times 5} = Q_{5\times 5} \begin{bmatrix} \Delta^{-1} & 0 & 0 & 0 & 0 \\ 0 & \Delta^{-1} & 0 & 0 & 0 \\ 0 & 0 & \Delta^{-2} & 0 & 0 \\ 0 & 0 & 0 & \Delta^{-2} & 0 \\ 0 & 0 & 0 & 0 & \Delta^{-2} \end{bmatrix} = \begin{bmatrix} \phi(y^1/\Delta)^\top \\ \phi(y^2/\Delta)^\top \\ \vdots \\ \phi(y^5/\Delta)^\top \end{bmatrix}.$$

Our goal is compute $\hat{Q}_{5\times 5}$ such that its inverse has the smallest possible norm:

$$\min \ \|(\hat{Q}_{5\times 5})^{-1}\|_2 \quad \Longleftrightarrow \quad \min \ \Lambda.$$

IDEA: Factorize the matrix $\hat{Q}_{5\times 5}$ by rows, changing the points in Y, if necessary, so that the pivots become as large as possible.

Ensuring well-poisedness (continued)

Consider a scaled version of $Q_{5\times 5}$:

$$\hat{Q}_{5\times 5} = Q_{5\times 5} \begin{bmatrix} \Delta^{-1} & 0 & 0 & 0 & 0 \\ 0 & \Delta^{-1} & 0 & 0 & 0 \\ 0 & 0 & \Delta^{-2} & 0 & 0 \\ 0 & 0 & 0 & \Delta^{-2} & 0 \\ 0 & 0 & 0 & 0 & \Delta^{-2} \end{bmatrix} = \begin{bmatrix} \phi(y^1/\Delta)^\top \\ \phi(y^2/\Delta)^\top \\ \vdots \\ \phi(y^5/\Delta)^\top \end{bmatrix}.$$

Our goal is compute $\hat{Q}_{5\times 5}$ such that its inverse has the smallest possible norm:

$$\min \; \|(\hat{Q}_{5\times 5})^{-1}\|_2 \quad \Longleftrightarrow \quad \min \; \Lambda.$$

IDEA: Factorize the matrix $\hat{Q}_{5\times 5}$ by rows, changing the points in Y, if necessary, so that the pivots become as large as possible.

Ensuring well-poisedness (continued)

Before eliminating the fourth row:

$$\longrightarrow \begin{bmatrix} \geq \xi & * & * & * & * \\ 0 & \geq \xi & * & * & * \\ 0 & 0 & \geq \xi & * & * \\ y_1^4/\Delta & y_2^4/\Delta & \frac{1}{2}(y_1^4/\Delta)^2 & \frac{1}{2}(y_2^4/\Delta)^2 & (y_1^4/\Delta)(y_2^4/\Delta) \\ y_1^5/\Delta & y_2^5/\Delta & \frac{1}{2}(y_1^5/\Delta)^2 & \frac{1}{2}(y_2^5/\Delta)^2 & (y_1^5/\Delta)(y_2^5/\Delta) \end{bmatrix}.$$

Fact 1: After factorization the element in the 4 × 4 position will be of the form:

$$v^\top \phi(y^4/\Delta)$$

where v, with $\|v\|_\infty \geq 1$, depends on the first 3 factorized rows.
Fact 2: However, for any $\|v\|_\infty \geq 1$, we can prove that: $\max_{y \in B(1)} |v^\top \phi(y)| \geq 1/4$. Thus, ξ can be as good as $1/4$.

Ensuring well-poisedness (continued)

Before eliminating the fourth row:

$$\longrightarrow \begin{bmatrix} \geq \xi & * & * & * & * \\ 0 & \geq \xi & * & * & * \\ 0 & 0 & \geq \xi & * & * \\ y_1^4/\Delta & y_2^4/\Delta & \frac{1}{2}(y_1^4/\Delta)^2 & \frac{1}{2}(y_2^4/\Delta)^2 & (y_1^4/\Delta)(y_2^4/\Delta) \\ y_1^5/\Delta & y_2^5/\Delta & \frac{1}{2}(y_1^5/\Delta)^2 & \frac{1}{2}(y_2^5/\Delta)^2 & (y_1^5/\Delta)(y_2^5/\Delta) \end{bmatrix}.$$

Fact 1: After factorization the element in the 4 × 4 position will be of the form:

$$v^\top \phi(y^4/\Delta)$$

where v, with $\|v\|_\infty \geq 1$, depends on the first 3 factorized rows.

Fact 2: However, for any $\|v\|_\infty \geq 1$, we can prove that: $\max_{y \in B(1)} |v^\top \phi(y)| \geq 1/4$. Thus, ξ can be as good as $1/4$.

Ensuring well-poisedness (continued)

Before eliminating the fourth row:

$$\longrightarrow \begin{bmatrix} \geq \xi & * & * & * & * \\ 0 & \geq \xi & * & * & * \\ 0 & 0 & \geq \xi & * & * \\ y_1^4/\Delta & y_2^4/\Delta & \frac{1}{2}(y_1^4/\Delta)^2 & \frac{1}{2}(y_2^4/\Delta)^2 & (y_1^4/\Delta)(y_2^4/\Delta) \\ y_1^5/\Delta & y_2^5/\Delta & \frac{1}{2}(y_1^5/\Delta)^2 & \frac{1}{2}(y_2^5/\Delta)^2 & (y_1^5/\Delta)(y_2^5/\Delta) \end{bmatrix}.$$

Fact 1: After factorization the element in the 4×4 position will be of the form:

$$v^\top \phi(y^4/\Delta)$$

where v, with $\|v\|_\infty \geq 1$, depends on the first 3 factorized rows.
Fact 2: However, for any $\|v\|_\infty \geq 1$, we can prove that:
$\max_{y \in B(1)} |v^\top \phi(y)| \geq 1/4$. Thus, ξ can be as good as $1/4$.

Ensuring well-poisedness (continued)

Before eliminating the fourth row:

$$\longrightarrow \begin{bmatrix} \geq \xi & * & * & * & * \\ 0 & \geq \xi & * & * & * \\ 0 & 0 & \geq \xi & * & * \\ y_1^4/\Delta & y_2^4/\Delta & \frac{1}{2}(y_1^4/\Delta)^2 & \frac{1}{2}(y_2^4/\Delta)^2 & (y_1^4/\Delta)(y_2^4/\Delta) \\ y_1^5/\Delta & y_2^5/\Delta & \frac{1}{2}(y_1^5/\Delta)^2 & \frac{1}{2}(y_2^5/\Delta)^2 & (y_1^5/\Delta)(y_2^5/\Delta) \end{bmatrix}.$$

Fact 1: After factorization the element in the 4 × 4 position will be of the form:

$$v^\top \phi(y^4/\Delta)$$

where v, with $\|v\|_\infty \geq 1$, depends on the first 3 factorized rows.
Fact 2: However, for any $\|v\|_\infty \geq 1$, we can prove that: $\max_{y \in B(1)} |v^\top \phi(y)| \geq 1/4$. Thus, ξ can be as good as $1/4$.

Ensuring well-poisedness (continued)

NOTES:

- The algorithm computes a set $Y \subset B(\Delta)$ of $p{=}(n+1)(n+2)/2$ points centered at $y^0 = 0$ for which the pivots of the GE of $\hat{Q}_{p-1 \times p-1}$ satisfy

$$|\xi_{ii}| \;\geq\; \frac{1}{4}, \;\; i = 1, \ldots, p-1.$$

- The effort required by the algorithm for the GE is $\mathcal{O}((n^2)^3) = \mathcal{O}(n^6)$ flops.
- In the worst case the algorithm requires, moreover, the minimization of $2(n-1)$ linear functions and $2(p-1-n)$ quadratic functions in a ball of radius 1.
- The algorithm can be applied to the linear case $(1/4 \rightarrow 1)$ or to any linear-quadratic setting, e.g. $\phi(x) \;=\; [\, x_1 \;\cdots\; x_n \; x_1^2/2 \;\cdots\; x_n^2/2 \,]^\top$.

Ensuring well-poisedness (continued)

NOTES:

- The algorithm computes a set $Y \subset B(\Delta)$ of $p{=}(n+1)(n+2)/2$ points centered at $y^0 = 0$ for which the pivots of the GE of $\hat{Q}_{p-1 \times p-1}$ satisfy

$$|\xi_{ii}| \;\geq\; \frac{1}{4}, \;\; i = 1, \ldots, p-1.$$

- The effort required by the algorithm for the GE is $\mathcal{O}((n^2)^3) = \mathcal{O}(n^6)$ flops.
- In the worst case the algorithm requires, moreover, the minimization of $2(n-1)$ linear functions and $2(p-1-n)$ quadratic functions in a ball of radius 1.
- The algorithm can be applied to the linear case $(1/4 \rightarrow 1)$ or to any linear-quadratic setting, e.g. $\phi(x) \;=\; [\, x_1 \;\cdots\; x_n \; x_1^2/2 \;\cdots\; x_n^2/2 \,]^\top$.

Ensuring well-poisedness (continued)

NOTES:

- The algorithm computes a set $Y \subset B(\Delta)$ of $p = (n+1)(n+2)/2$ points centered at $y^0 = 0$ for which the pivots of the GE of $\hat{Q}_{p-1 \times p-1}$ satisfy

$$|\xi_{ii}| \;\geq\; \frac{1}{4}, \; i = 1, \ldots, p-1.$$

- The effort required by the algorithm for the GE is $\mathcal{O}((n^2)^3) = \mathcal{O}(n^6)$ flops.
- In the worst case the algorithm requires, moreover, the minimization of $2(n-1)$ linear functions and $2(p-1-n)$ quadratic functions in a ball of radius 1.
- The algorithm can be applied to the linear case $(1/4 \rightarrow 1)$ or to any linear-quadratic setting, e.g. $\phi(x) \; = \; [\, x_1 \, \cdots \, x_n \; x_1^2/2 \, \cdots \, x_n^2/2 \,]^\top$.

Ensuring well-poisedness (continued)

NOTES:

- The algorithm computes a set $Y \subset B(\Delta)$ of $p{=}(n+1)(n+2)/2$ points centered at $y^0 = 0$ for which the pivots of the GE of $\hat{Q}_{p-1 \times p-1}$ satisfy

$$|\xi_{ii}| \;\geq\; \frac{1}{4}, \;\; i = 1, \ldots, p-1.$$

- The effort required by the algorithm for the GE is $\mathcal{O}((n^2)^3) = \mathcal{O}(n^6)$ flops.
- In the worst case the algorithm requires, moreover, the minimization of $2(n-1)$ linear functions and $2(p-1-n)$ quadratic functions in a ball of radius 1.
- The algorithm can be applied to the linear case $(1/4 \rightarrow 1)$ or to any linear-quadratic setting, e.g. $\phi(x) \;=\; [\, x_1 \;\cdots\; x_n \; x_1^2/2 \;\cdots\; x_n^2/2 \,]^\top$.

Other error estimates

Error estimates (in an geometry setting inappropriate to DFO):

- Ciarlet and Raviart, *General Lagrange and Hermite interpolation in $\mathbb{R}^n$ with applications to finite element methods*, Arch. Rational Mech. Anal., 46 (1972) 177-199.

Derivative bounds involve constants exponentially growing and dependent on Y.

Other approaches, more suitable for DFO, but with no estimate for the error in the derivatives and without a rigorous control of the geometry:

- Sauer and Xu, *On multivariate Lagrange interpolation*, Math. Comp., 64 (1995) 1147-1170 — Newton polynomials.
- S. Waldron, Multipoint Taylor formulae, Numer. Math. 80 (1998).
- Powell, *On the Lagrange functions of quadratic models that are defined by interpolation*, Optim. Methods Softw., 16 (2001) 289-309 — Lagrange polynomials.

Other error estimates

Error estimates (in an geometry setting inappropriate to DFO):

- Ciarlet and Raviart, *General Lagrange and Hermite interpolation in $\mathbb{R}^n$ with applications to finite element methods*, Arch. Rational Mech. Anal., 46 (1972) 177-199.

Derivative bounds involve constants exponentially growing and dependent on Y.

Other approaches, more suitable for DFO, but with no estimate for the error in the derivatives and without a rigorous control of the geometry:

- Sauer and Xu, *On multivariate Lagrange interpolation*, Math. Comp., 64 (1995) 1147-1170 — Newton polynomials.
- S. Waldron, Multipoint Taylor formulae, Numer. Math. 80 (1998).
- Powell, *On the Lagrange functions of quadratic models that are defined by interpolation*, Optim. Methods Softw., 16 (2001) 289-309 — Lagrange polynomials.

Other error estimates

Error estimates (in an geometry setting inappropriate to DFO):

- Ciarlet and Raviart, *General Lagrange and Hermite interpolation in $\mathbb{R}^n$ with applications to finite element methods*, Arch. Rational Mech. Anal., 46 (1972) 177-199.

Derivative bounds involve constants exponentially growing and dependent on Y.

Other approaches, more suitable for DFO, but with no estimate for the error in the derivatives and without a rigorous control of the geometry:

- Sauer and Xu, *On multivariate Lagrange interpolation*, Math. Comp., 64 (1995) 1147-1170 — Newton polynomials.
- S. Waldron, Multipoint Taylor formulae, Numer. Math. 80 (1998).
- Powell, *On the Lagrange functions of quadratic models that are defined by interpolation*, Optim. Methods Softw., 16 (2001) 289-309 — Lagrange polynomials.

Our work

- Conn, Gould, and Toint, **Trust-Region Methods**, SIAM Publications, 2000.
- Conn and Toint, *An algorithm using quadratic interpolation for unconstrained derivative free optimization*, in Nonlinear Optimization and Applications, Plenum, 22-47, 1996.
- Conn, Scheinberg, and Toint, *On the convergence of derivative free methods for unconstrained optimization*, in Approximation Theory and Optimization, CUP, 83-108, 1997.
- Conn, Scheinberg, and Toint, *Recent progress in unconstrained nonlinear optimization without derivatives*, MP, 79(3):397-414, 1997.

Our work (continued)

- Conn, Scheinberg, and Vicente, *Geometry of Sample Sets in Derivative free Optimization. Part I: Polynomial Interpolation*, MP, appeared electronically, 2006.

- Conn, Scheinberg, and Vicente, *Geometry of Sample Sets in Derivative free Optimization. Part II: Regression and Underdetermined Interpolation*, 2005.

- Conn, Kelley, Scheinberg, and Vicente, *Derivative-free optimization with variable geometry III: extensions of simplex derivatives and implicit filtering*, paper under preparation.

- Scheinberg *DFO Software (Open Source)*, http://www.coin-or.org/.

Other Approaches

Pattern Search Methods

Box (57), Campey-Nickols (61), Hooke-Jeeves (61), Spendley-Hext-Himsworth (62), Nelder-Meade (65), Dixon (73) [ACSIM], Dennis-Torczon (91)[PDS], Buckley (94), Wright (96), Torczon (97) Dennis-Torczon(97), Alexandrov et. al. (98), Audet (99), Audet-Dennis (00), Coope-Price (00), Hough et. al. (01), Coope et. al. (02), Kolda-Torczon (03) ...

Linesearch Methods

Powell (64), [VA04,VA24], Brent (73) [PRAXIS],Toint-Callier (77), Lucidi-Sciandrone (96, 02), Custódio-Vicente (04) ...

Finite Difference Gradient & Quasi-Newton

Powell (65, 70), Stewart (67), Gill-Murray-Wright (81), Dennis-Schnabel (83)

Other Approaches

Pattern Search Methods

Box (57), Campey-Nickols (61), Hooke-Jeeves (61), Spendley-Hext-Himsworth (62), Nelder-Meade (65), Dixon (73) [ACSIM], Dennis-Torczon (91)[PDS], Buckley (94), Wright (96), Torczon (97) Dennis-Torczon(97), Alexandrov et. al. (98), Audet (99), Audet-Dennis (00), Coope-Price (00), Hough et. al. (01), Coope et. al. (02), Kolda-Torczon (03) ...

Linesearch Methods

Powell (64), [VA04,VA24], Brent (73) [PRAXIS],Toint-Callier (77), Lucidi-Sciandrone (96, 02), Custódio-Vicente (04) ...

Finite Difference Gradient & Quasi-Newton

Powell (65, 70), Stewart (67), Gill-Murray-Wright (81), Dennis-Schnabel (83)

Other Approaches

Pattern Search Methods

Box (57), Campey-Nickols (61), Hooke-Jeeves (61), Spendley-Hext-Himsworth (62), Nelder-Meade (65), Dixon (73) [ACSIM], Dennis-Torczon (91)[PDS], Buckley (94), Wright (96), Torczon (97) Dennis-Torczon(97), Alexandrov et. al. (98), Audet (99), Audet-Dennis (00), Coope-Price (00), Hough et. al. (01), Coope et. al. (02), Kolda-Torczon (03) ...

Linesearch Methods

Powell (64), [VA04,VA24], Brent (73) [PRAXIS],Toint-Callier (77), Lucidi-Sciandrone (96, 02), Custódio-Vicente (04) ...

Finite Difference Gradient & Quasi-Newton

Powell (65, 70), Stewart (67), Gill-Murray-Wright (81), Dennis-Schnabel (83)

References (continued)

Interpolation Methods

Powell(94)(02)[UOBYQA], Colson (03), Toint-Colson (04)

[Excellent Survey Article for Pattern Search Methods]

Kolda Lewis and Torczon

Pattern Search Survey *Optimization by Direct Search:New Perspectives on Some Classical and Modern Methods,*

SIAM Rev., 45(3):385-482, 2003.

References (continued)

Interpolation Methods

Powell(94)(02)[UOBYQA], Colson (03), Toint-Colson (04)

[Excellent Survey Article for Pattern Search Methods]

Kolda Lewis and Torczon

Pattern Search Survey *Optimization by Direct Search:New Perspectives on Some Classical and Modern Methods*,

SIAM Rev., 45(3):385-482, 2003.

References (continued)

Interpolation Methods

Powell(94)(02)[UOBYQA], Colson (03), Toint-Colson (04)

[Excellent Survey Article for Pattern Search Methods]

Kolda Lewis and Torczon

Pattern Search Survey *Optimization by Direct Search:New Perspectives on Some Classical and Modern Methods*,

SIAM Rev., 45(3):385-482, 2003.

References (continued)

Interpolation Methods

Powell(94)(02)[UOBYQA], Colson (03), Toint-Colson (04)

[Excellent Survey Article for Pattern Search Methods]

Kolda Lewis and Torczon
Pattern Search Survey *Optimization by Direct Search:New Perspectives on Some Classical and Modern Methods*,
SIAM Rev., 45(3):385-482, 2003.