Computation of Sparse Low Degree Interpolating Polynomials and their Application to DFO

Afonso Bandeira[1] Katya Scheinberg[2] Luis Nunes Vicente[1]

[1]University of Coimbra [2]Columbia University

EUROPT 2010, Aveiro, July 9, 2010

Presentation outline

1. Motivation and general concepts
2. Sparse reconstruction
3. Sparse Hessian construction
4. A practical interpolation-based trust-region method for DFO
5. Concluding remarks

Presentation outline

Derivative-free optimization

Solve:

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

Derivative-free optimization

Solve:

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

- function with unknown derivatives

Derivative-free optimization

Solve:

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

- function with unknown derivatives
- expensive function evaluations, possibly subject to noise

Derivative-free optimization

Solve:

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

- function with unknown derivatives
- expensive function evaluations, possibly subject to noise
- unpractical to compute approximations to derivatives

Model-based trust-region methods

- One typically minimizes a model m in a trust region $B_p(x;\Delta)$:

Trust-region subproblem

$$\min_{u \in B_p(x;\Delta)} m(u)$$

Model-based trust-region methods

- One typically minimizes a model m in a trust region $B_p(x;\Delta)$:

Trust-region subproblem

$$\min_{u \in B_p(x;\Delta)} m(u)$$

- m simple enough to be easy to minimize.

Model-based trust-region methods

- One typically minimizes a model m in a trust region $B_p(x;\Delta)$:

Trust-region subproblem

$$\min_{u \in B_p(x;\Delta)} m(u)$$

- m simple enough to be easy to minimize.
- m complex enough to well approximate f
(e.g., linear models do not capture the curvature of f).

Fully quadratic models

- Impossible to use Taylor models. — The derivatives are unknown.

Fully quadratic models

- Impossible to use Taylor models. — The derivatives are unknown.
- We still want models 'as good' as quadratic Taylor models.

Fully quadratic models

- Impossible to use Taylor models. — The derivatives are unknown.
- We still want models 'as good' as quadratic Taylor models.

Definition

m is a fully quadratic model for f on $B_p(x;\Delta)$ if, $\forall u \in B_p(x;\Delta)$,

- $$\|\nabla^2 f(u) - \nabla^2 m(u)\|_2 \ \le\ \kappa_{eh}\,\Delta,$$
- $$\|\nabla f(u) - \nabla m(u)\|_2 \ \le\ \kappa_{eg}\,\Delta^2,$$
- $$|f(u) - m(u)| \ \le\ \kappa_{ef}\,\Delta^3.$$

Fully quadratic models

- Impossible to use Taylor models. — The derivatives are unknown.
- We still want models 'as good' as quadratic Taylor models.

Definition

m is a fully quadratic model for f on $B_p(x;\Delta)$ if, $\forall u \in B_p(x;\Delta)$,

- $\|\nabla^2 f(u) - \nabla^2 m(u)\|_2 \leq \kappa_{eh}\, \Delta,$
- $\|\nabla f(u) - \nabla m(u)\|_2 \leq \kappa_{eg}\, \Delta^2,$
- $|f(u) - m(u)| \leq \kappa_{ef}\, \Delta^3.$

- How to construct fully quadratic models ($\forall \Delta > 0$)?

(Quadratic) interpolation

Let $\mathcal{P} = \mathrm{span}\{\phi_1, ..., \phi_N\}$ be a finite dimensional functional space (in $\mathbb{R}^n$).

(Quadratic) interpolation

Let $\mathcal{P} = \mathrm{span}\{\phi_1, ..., \phi_N\}$ be a finite dimensional functional space (in $\mathbb{R}^n$).

$m = \sum_{j=1}^{N} \alpha_j \phi_j$ interpolates f at $W = \{w_1, ..., w_k\}$ iff

$$M(\phi, W)\alpha = f(W).$$

(Quadratic) interpolation

Let $\mathcal{P} = \mathrm{span}\{\phi_1, ..., \phi_N\}$ be a finite dimensional functional space (in $\mathbb{R}^n$).

$m = \sum_{j=1}^{N} \alpha_j \phi_j$ interpolates f at $W = \{w_1, ..., w_k\}$ iff

$$M(\phi, W)\alpha = f(W).$$

One important example:

$$\bar{\phi} = \left\{1, u_1, ..., u_n, \frac{1}{2}u_1^2, ..., \frac{1}{2}u_n^2, u_1 u_2, ..., u_{n-1}u_n\right\}.$$

(Quadratic) interpolation

Let $\mathcal{P} = \mathrm{span}\{\phi_1, ..., \phi_N\}$ be a finite dimensional functional space (in $\mathbb{R}^n$).

$m = \sum_{j=1}^{N} \alpha_j \phi_j$ interpolates f at $W = \{w_1, ..., w_k\}$ iff

$$M(\phi, W)\alpha = f(W).$$

One important example:

$$\bar{\phi} = \left\{1, u_1, ..., u_n, \frac{1}{2}u_1^2, ..., \frac{1}{2}u_n^2, u_1u_2, ..., u_{n-1}u_n\right\}.$$

- Issue: For $M(\phi, W)$ to be nonsingular ones needs $N = (n+2)(n+1)/2$ evaluations of f (often too expensive).

(Quadratic) interpolation

Let $\mathcal{P} = \mathrm{span}\{\phi_1, ..., \phi_N\}$ be a finite dimensional functional space (in $\mathbb{R}^n$).

$m = \sum_{j=1}^{N} \alpha_j \phi_j$ interpolates f at $W = \{w_1, ..., w_k\}$ iff

$$M(\phi, W)\alpha = f(W).$$

One important example:

$$\bar{\phi} = \left\{1, u_1, ..., u_n, \frac{1}{2}u_1^2, ..., \frac{1}{2}u_n^2, u_1 u_2, ..., u_{n-1}u_n\right\}.$$

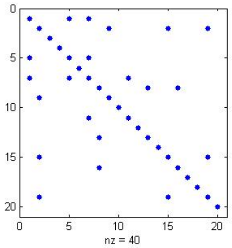
- Issue: For $M(\phi, W)$ to be nonsingular ones needs $N = (n+2)(n+1)/2$ evaluations of f (often too expensive).
- One possible Fix: Underdetermined Interpolation.

Sparsity on the Hessian

- Many pairs of variables have no correlation, leading to zero second order partial derivatives in f:

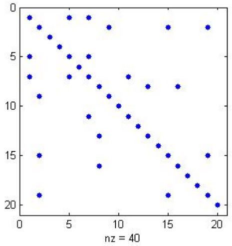
Sparsity on the Hessian

- Many pairs of variables have no correlation, leading to zero second order partial derivatives in f:



Sparsity on the Hessian

- Many pairs of variables have no correlation, leading to zero second order partial derivatives in f:



- Thus, the Hessian $\nabla^2 m(x = 0)$ of the model (i.e., the vector $\alpha_{\bar{\phi}}$ in the basis $\bar{\phi}$) should be sparse.

Our main question

Is it possible to build fully quadratic models by quadratic underdetermined interpolation (i.e., using less than $N = (n+2)(n+1)/2 = \mathcal{O}(n^2)$ points) in the sparse case?

Presentation outline

Compressed sensing — sparse recovery

- Objective: Find sparse α subject to a highly underdetermined linear system $M\alpha = f$.

Compressed sensing — sparse recovery

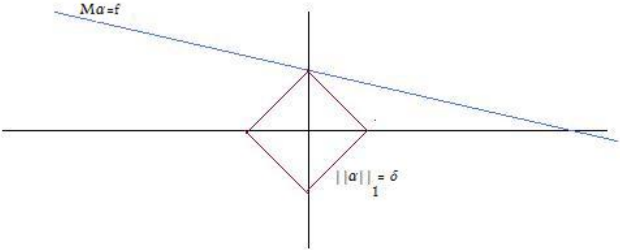
- Objective: Find sparse α subject to a highly underdetermined linear system $M\alpha = f$.
- $\begin{cases} \min & \|\alpha\|_0 = |\operatorname{supp}(\alpha)| \\ \text{s.t.} & M\alpha = f \end{cases}$ is NP-Hard.

Compressed sensing — sparse recovery

- Objective: Find sparse α subject to a highly underdetermined linear system $M\alpha = f$.
- $\begin{cases} \min & \|\alpha\|_0 \quad = |\operatorname{supp}(\alpha)| \\ \text{s.t.} & M\alpha = f \end{cases}$ is NP-Hard.
- $\begin{cases} \min & \|\alpha\|_1 \\ \text{s.t.} & M\alpha = f \end{cases}$ often recovers sparse solutions.

Compressed sensing — sparse recovery

- Objective: Find sparse α subject to a highly underdetermined linear system $M\alpha = f$.
- $\begin{cases} \min & \|\alpha\|_0 = |\operatorname{supp}(\alpha)| \\ \text{s.t.} & M\alpha = f \end{cases}$ is NP-Hard.
- $\begin{cases} \min & \|\alpha\|_1 \\ \text{s.t.} & M\alpha = f \end{cases}$ often recovers sparse solutions.



Restricted isometry property

Definition (RIP)

The RIP Constant of order s of M $(k \times N)$ is the smallest δ_s such that

$$(1-\delta_s)\|\alpha\|_2^2 \leq \|M\alpha\|_2^2 \leq (1+\delta_s)\|\alpha\|_2^2$$

for all $s-$sparse α ($\|\alpha\|_0 \leq s$).

Restricted isometry property

Definition (RIP)

The RIP Constant of order s of M $(k \times N)$ is the smallest δ_s such that

$$(1-\delta_s)\|\alpha\|_2^2 \leq \|M\alpha\|_2^2 \leq (1+\delta_s)\|\alpha\|_2^2$$

for all $s-$sparse α ($\|\alpha\|_0 \leq s$).

Theorem (Candès, Tau, 2005, 2006)

If α is $s-$sparse and $2\delta_{2s}+\delta_s < 1$ then α is recovered by ℓ_1-minimization.

Restricted isometry property

Definition (RIP)

The RIP Constant of order s of M $(k \times N)$ is the smallest δ_s such that

$$(1-\delta_s)\|\alpha\|_2^2 \leq \|M\alpha\|_2^2 \leq (1+\delta_s)\|\alpha\|_2^2$$

for all $s-$sparse α ($\|\alpha\|_0 \leq s$).

Theorem (Candès, Tau, 2005, 2006)

If α is $s-$sparse and $2\delta_{2s} + \delta_s < 1$ then α is recovered by ℓ_1-minimization.

Similar results hold for noisy measurements

$$M\alpha = f + \epsilon.$$

Random matrices

- It is hard to find deterministic matrices that satisfy the RIP for large s.

Random matrices

- It is hard to find deterministic matrices that satisfy the RIP for large s.

- Using Random Matrices it is possible to find RIP Matrices for

$$k = \mathcal{O}(s \log N).$$

  - Gaussian ensembles.
  - Bernoulli ensembles.
  - Uniformly chosen subsets of discrete Fourier transform.

Presentation outline

Bounded orthonormal expansions

Question

How to find a basis ϕ and a sample set W such that $M(\phi, W)$ satisfies the RIP?

Bounded orthonormal expansions

Question

How to find a basis ϕ *and a sample set* W *such that* $M(\phi, W)$ *satisfies the RIP?*

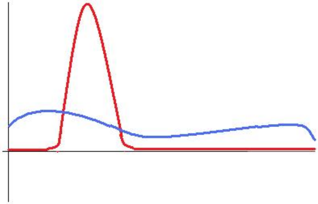
- Orthogonal bases.

Bounded orthonormal expansions

Question

How to find a basis ϕ and a sample set W such that $M(\phi, W)$ satisfies the RIP?

- Orthogonal bases.
- No localized functions.

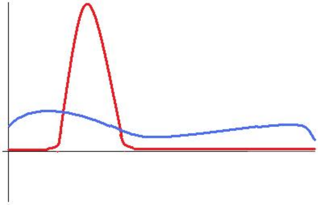


Bounded orthonormal expansions

Question

How to find a basis ϕ and a sample set W such that $M(\phi, W)$ satisfies the RIP?

- Orthogonal bases.
- No localized functions.



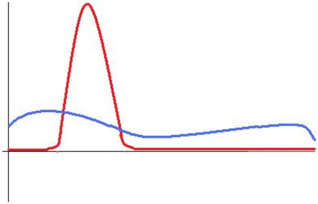
- $\frac{\|\phi_i\|_{L^\infty}}{\|\phi_i\|_{L^2}}$ should be uniformly bounded (by K).

Bounded orthonormal expansions

Question

How to find a basis ϕ and a sample set W such that $M(\phi, W)$ satisfies the RIP?

- Orthogonal bases.
- No localized functions.



- $\frac{\|\phi_i\|_{L^\infty}}{\|\phi_i\|_{L^2}}$ should be uniformly bounded (by K).
- W will be a random sample set.

Sparse orthonormal bounded expansion recovery

Theorem (Rauhut, 2010)

If
- ϕ *is orthonormal in a probably measure* μ *and* $\|\phi_i\|_{L^\infty} \leq K$*.*
- *each point of* W *is drawn independently according to* μ*.*
- $\frac{k}{\log k} \geq c_1 K^2 s (\log s)^2 \log N$*.*

Sparse orthonormal bounded expansion recovery

Theorem (Rauhut, 2010)

If
- ϕ *is orthonormal in a probably measure* μ *and* $\|\phi_i\|_{L^\infty} \leq K$.
- *each point of* W *is drawn independently according to* μ.
- $\frac{k}{\log k} \geq c_1 K^2 s (\log s)^2 \log N$.

Then, with high probability, for every $s-$*sparse vector* α*:*

Given noisy samples $f = M(\phi, W)\alpha + \epsilon$ *with* $\|\epsilon\|_2 \leq \eta$, *let* α^* *be the solution of*

$$\min \|\alpha\|_1 \quad \text{s.t.} \quad \|M(\phi, W)\alpha - f\|_2 \leq \eta.$$

Then,

$$\|\alpha - \alpha^*\|_2 \leq \frac{d}{\sqrt{k}}\eta.$$

A suitable basis for quadratics in $B_\infty(0;\Delta)$

Proposition

The following basis ψ for quadratics in $B_\infty(0;\Delta)$ is orthonormal and satisfies $\|\psi_\iota\|_{L^\infty} \leq 3$.

$$\begin{cases} \psi_0(u) & = & 1 \\ \psi_{1,i}(u) & = & \frac{\sqrt{3}}{\Delta} u_i \\ \psi_{2,ij}(u) & = & \frac{3}{\Delta^2} u_i u_j \\ \psi_{2,i}(u) & = & \frac{3\sqrt{5}}{2} \frac{1}{\Delta^2} u_i^2 - \frac{\sqrt{5}}{2}. \end{cases}$$

A suitable basis for quadratics in $B_\infty(0;\Delta)$

Proposition

The following basis ψ for quadratics in $B_\infty(0;\Delta)$ is orthonormal and satisfies $\|\psi_\iota\|_{L^\infty} \leq 3$.

$$\begin{cases} \psi_0(u) & = & 1 \\ \psi_{1,i}(u) & = & \frac{\sqrt{3}}{\Delta} u_i \\ \psi_{2,ij}(u) & = & \frac{3}{\Delta^2} u_i u_j \\ \psi_{2,i}(u) & = & \frac{3\sqrt{5}}{2}\frac{1}{\Delta^2} u_i^2 - \frac{\sqrt{5}}{2}. \end{cases}$$

- ψ is very similar to the canonical basis:
  - preserves the sparsity of the Hessian (at 0).
  - changes fast to the canonical basis (and vice-versa).

Hessian recovery

Theorem (Main Theorem)

If

- *The Hessian of f at 0 is $s-$sparse.*
- *W is a random sample set chosen with respect to the uniform measure in $B_\infty(0;\infty)$.*
- $\frac{k}{\log k} \geq 9c_1(s+n+1)\log^2(s+n+1)\log N$.

Hessian recovery

Theorem (Main Theorem)

If

- *The Hessian of* f *at* 0 *is* $s-$*sparse.*
- W *is a random sample set chosen with respect to the uniform measure in* $B_\infty(0;\infty)$*.*
- $\frac{k}{\log k} \geq 9c_1(s+n+1)\log^2(s+n+1)\log N$*.*

Then, with high probability, the quadratic

$$q^* = \sum \alpha_\iota^* \psi_\iota$$

obtained by solving the noisy ℓ_1*-minimization problem is a fully quadratic model for* f *(with error constants not depending on* Δ*).*

Remarks

- We recover both the function and the sparsity structure.

Remarks

- We recover both the function and the sparsity structure.
- We are able to construct fully quadratic models with $\mathcal{O}(n \log^4 n)$ points when the number of non-zeros of the Hessian is $s = \mathcal{O}(n)$.

Remarks

- We recover both the function and the sparsity structure.
- We are able to construct fully quadratic models with $\mathcal{O}(n \log^4 n)$ points when the number of non-zeros of the Hessian is $s = \mathcal{O}(n)$.
- However, the Theorem only provides motivation because, in a practical approach we:

Remarks

- We recover both the function and the sparsity structure.
- We are able to construct fully quadratic models with $\mathcal{O}(n \log^4 n)$ points when the number of non-zeros of the Hessian is $s = \mathcal{O}(n)$.
- However, the Theorem only provides motivation because, in a practical approach we:
  - solve

$$\begin{array}{ll} \min & \|\alpha_Q\|_1 \\ \text{s.t.} & M(\bar{\phi}_L, W)\alpha_L + M(\bar{\phi}_Q, W)\alpha_Q = f(W), \end{array}$$

Remarks

- We recover both the function and the sparsity structure.
- We are able to construct fully quadratic models with $\mathcal{O}(n \log^4 n)$ points when the number of non-zeros of the Hessian is $s = \mathcal{O}(n)$.
- However, the Theorem only provides motivation because, in a practical approach we:
  - solve

$$\begin{array}{ll} \min & \|\alpha_Q\|_1 \\ \text{s. t.} & M(\bar{\phi}_L, W)\alpha_L + M(\bar{\phi}_Q, W)\alpha_Q = f(W), \end{array}$$

  - deal with small n (from the DFO setting) and the bound we obtain is asymptotical,

Remarks

- We recover both the function and the sparsity structure.
- We are able to construct fully quadratic models with $\mathcal{O}(n \log^4 n)$ points when the number of non-zeros of the Hessian is $s = \mathcal{O}(n)$.
- However, the Theorem only provides motivation because, in a practical approach we:
  - solve
$$\begin{array}{ll} \min & \|\alpha_Q\|_1 \\ \text{s. t.} & M(\bar{\phi}_L, W)\alpha_L + M(\bar{\phi}_Q, W)\alpha_Q = f(W), \end{array}$$
  - deal with small n (from the DFO setting) and the bound we obtain is asymptotical,
  - use determinist sampling.

Presentation outline

Algorithm: Part I

Initialize the sample set Y with $2n+1$ points and evaluate f there.

Then, repeat until the stopping criterion (small trust-region radius or model gradient) is achieved:

Algorithm: Part I

Initialize the sample set Y with $2n+1$ points and evaluate f there.

Then, repeat until the stopping criterion (small trust-region radius or model gradient) is achieved:

- Model building: find m_k in $B(x_k; \Delta_k)$.
  - If there are enough points use determined quadratic interpolation.
  - Otherwise use ℓ_1 $(p=1)$ or Frobenius $(p=2)$ minimum norm quadratic interpolation:

$$\begin{array}{ll} \min & \frac{1}{p}\|\alpha_Q\|_p^p \\ \text{s.t.} & M(\bar{\phi}_L, W)\alpha_L + M(\bar{\phi}_Q, W)\alpha_Q = f(W). \end{array}$$

Algorithm: Part I

Initialize the sample set Y with $2n+1$ points and evaluate f there.

Then, repeat until the stopping criterion (small trust-region radius or model gradient) is achieved:

- Model building: find m_k in $B(x_k;\Delta_k)$.
  - If there are enough points use determined quadratic interpolation.
  - Otherwise use ℓ_1 $(p=1)$ or Frobenius $(p=2)$ minimum norm quadratic interpolation:

$$\begin{array}{ll} \min & \frac{1}{p}\|\alpha_Q\|_p^p \\ \text{s.t.} & M(\bar{\phi}_L, W)\alpha_L + M(\bar{\phi}_Q, W)\alpha_Q = f(W). \end{array}$$

- Step calculation: Find s_k by solving the trust-region subproblem

$$\min_{s \in B_2(0;\Delta_k)} m_k(x_k + s).$$

Algorithm: Part II (inspired by Fasano, Morales, and Nocedal, 2009)

Algorithm: Part II (inspired by Fasano, Morales, and Nocedal, 2009)

- Iterate and trust-region radius update: Compute

$$\rho_k = \frac{f(x_k) - f(x_k + s_k)}{m(x_k) - m(x_k + s_k)}.$$

Algorithm: Part II (inspired by Fasano, Morales, and Nocedal, 2009)

- Iterate and trust-region radius update: Compute

$$\rho_k = \frac{f(x_k) - f(x_k + s_k)}{m(x_k) - m(x_k + s_k)}.$$

  - if $\rho_k < \eta_0$, then $x_{k+1} = x_k$ and $\Delta_{k+1} < \Delta_k$ (unsuccessful).
  - if $\eta_0 \leq \rho_k < \eta_1$, then $x_{k+1} = x_k + s_k$ and $\Delta_{k+1} < \Delta_k$ (acceptable).
  - if $\rho_k \geq \eta_1$, then $x_{k+1} = x_k + s_k$ and $\Delta_{k+1} \geq \Delta_k$ (successful).

Algorithm: Part II (inspired by Fasano, Morales, and Nocedal, 2009)

- Iterate and trust-region radius update: Compute

$$\rho_k = \frac{f(x_k) - f(x_k + s_k)}{m(x_k) - m(x_k + s_k)}.$$

  - if $\rho_k < \eta_0$, then $x_{k+1} = x_k$ and $\Delta_{k+1} < \Delta_k$ (unsuccessful).
  - if $\eta_0 \leq \rho_k < \eta_1$, then $x_{k+1} = x_k + s_k$ and $\Delta_{k+1} < \Delta_k$ (acceptable).
  - if $\rho_k \geq \eta_1$, then $x_{k+1} = x_k + s_k$ and $\Delta_{k+1} \geq \Delta_k$ (successful).

- Sample set update:
  - if $|Y_k| \leq |Y_{max}|$, set $Y_{k+1} = Y_k \cup \{x_k + s_k\}$.
  - otherwise set $y_{out} = \text{argmax} \|y - x_{k+1}\|_2$ and $Y_{k+1} = Y_k \cup \{x_k + s_k\} \backslash y_{out}$.

Algorithm: Part II (inspired by Fasano, Morales, and Nocedal, 2009)

- Iterate and trust-region radius update: Compute

$$\rho_k = \frac{f(x_k) - f(x_k + s_k)}{m(x_k) - m(x_k + s_k)}.$$

  - if $\rho_k < \eta_0$, then $x_{k+1} = x_k$ and $\Delta_{k+1} < \Delta_k$ (unsuccessful).
  - if $\eta_0 \leq \rho_k < \eta_1$, then $x_{k+1} = x_k + s_k$ and $\Delta_{k+1} < \Delta_k$ (acceptable).
  - if $\rho_k \geq \eta_1$, then $x_{k+1} = x_k + s_k$ and $\Delta_{k+1} \geq \Delta_k$ (successful).

- Sample set update:
  - if $|Y_k| \leq |Y_{max}|$, set $Y_{k+1} = Y_k \cup \{x_k + s_k\}$.
  - otherwise set $y_{out} = \text{argmax} \|y - x_{k+1}\|_2$ and $Y_{k+1} = Y_k \cup \{x_k + s_k\} \backslash y_{out}$.

- 'Criticality step': If the trust-radius radius is very small discard points far away from the trust region.

Performance profiles (accuracy of 10^{-4} in function values)

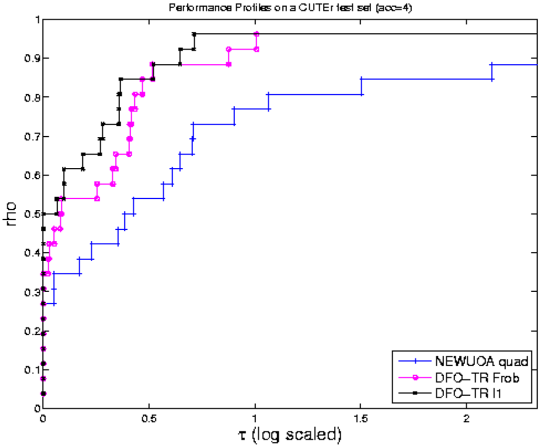


Figure: Performance profiles comparing DFO−TR (ℓ_1 and Frobenius) and NEWUOA (Powell) in a test set from CUTEr.

Performance profiles (accuracy of 10^{-6} in function values)

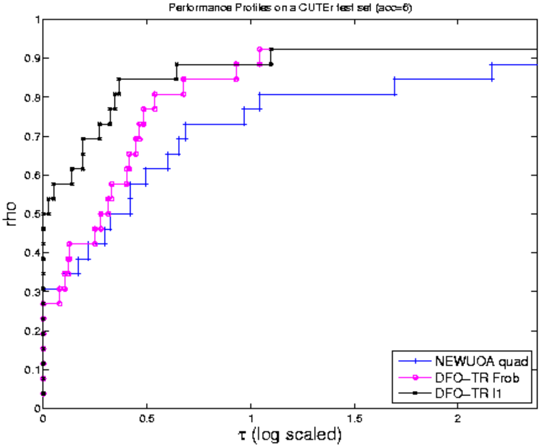


Figure: Performance profiles comparing DFO-TR (ℓ_1 and Frobenius) and NEWUOA (Powell) in a test set from CUTEr.

DFO-TR: ℓ_1 versus Frobenius

problem	DFO-TR Frob/l1	# f eval	f val	model ∇ norm
ARWHEAD	Frob	338	3.044e-07	3.627e-03
ARWHEAD	l1	218	9.168e-11	7.651e-07
BRYDN3D	Frob	41	0.000e+00	0.000e+00
BRYDN3D	l1	41	0.000e+00	0.000e+00
DQDRTIC	Frob	72	8.709e-11	6.300e+05
DQDRTIC	l1	45	8.693e-13	1.926e-06
EXTROSNB	Frob	1068	6.465e-02	3.886e+02
EXTROSNB	l1	2070	1.003e-02	6.750e-02
SROSENBR	Frob	456	2.157e-03	4.857e-02
SROSENBR	l1	297	1.168e-02	3.144e-01
WOODS	Frob	5000	1.902e-01	8.296e-01
WOODS	l1	5000	1.165e+01	1.118e+01

Table: A sample of test problems from CUTEr (Hessian sparse).

Presentation outline

Concluding remarks

- Compressed Sensing has been tightly connected to Optimization, since Optimization is a fundamental tool in CS. However, this work shows that CS can also be 'applied to' Optimization.

Concluding remarks

- Compressed Sensing has been tightly connected to Optimization, since Optimization is a fundamental tool in CS. However, this work shows that CS can also be 'applied to' Optimization.

- In a sparse scenario, we were able to construct fully quadratic models with samples of size $\mathcal{O}(n \log^4 n)$ instead of the classical $\mathcal{O}(n^2)$.

Concluding remarks

- Compressed Sensing has been tightly connected to Optimization, since Optimization is a fundamental tool in CS. However, this work shows that CS can also be 'applied to' Optimization.

- In a sparse scenario, we were able to construct fully quadratic models with samples of size $\mathcal{O}(n \log^4 n)$ instead of the classical $\mathcal{O}(n^2)$.

- We proposed a practical DFO method (using ℓ_1-minimization) that was able to outperform state-of-the-art methods in several numerical tests (in the already 'tough' DFO scenario where n is small).

Future work

- Gain understanding about the advantages of minimizing the ℓ_1-norm of the Hessian (α_Q) and not of the whole α. Numerical simulations show that such approach is advantageous.

Future work

- Gain understanding about the advantages of minimizing the ℓ_1-norm of the Hessian (α_Q) and not of the whole α. Numerical simulations show that such approach is advantageous.

- Improve the efficiency of the model ℓ_1-minimization, may be by finding a way to warmstart it.

Future work

- Gain understanding about the advantages of minimizing the ℓ_1-norm of the Hessian (α_Q) and not of the whole α. Numerical simulations show that such approach is advantageous.
- Improve the efficiency of the model ℓ_1-minimization, may be by finding a way to warmstart it.
- Study the convergence properties of possibly stochastic model-based trust-region methods.

References

A. Bandeira, *Computation of Sparse Low Degree Interpolating Polynomials and their Application to Derivative-Free Optimization*, Master Thesis, Dept. Mathematics, Univ. Coimbra, 2010.

A. Bandeira, K. Scheinberg, and L. N. Vicente, *Computation of sparse low degree interpolating polynomials and their application to derivative-free optimization*, in preparation, 2010.