4TE3/6TE3

Algorithms for

Continuous Optimization

DFO-Trust Region Interpolation Algorithm

Tamás TERLAKY
Computing and Software
McMaster University

Hamilton, January 2004

terlaky@mcmaster.ca Tel: 27780

Trust region framework

Steps of a trust region method

1. Given a current iterate build a good local approximation model (e.g., based on a second order Taylor series approximation).
2. Choose a neighborhood around the current iterate where the model is "trusted" to be accurate. Minimize the model in this neighborhood.
3. Determine if the step is successful by evaluating the true function at the new point and comparing the true reduction in value of the objective with the reduction predicted by the model.
4. If the step is successful, accept the new point as the next iterate and proceed (possibly, increasing the size of the trust region if the success is really significant).

   If the step is unsuccessful, reject the new point and reduce the size of the trust region.
5. Repeat until convergence.

Check the paper "qi-trust".

Interpolation of functions

by using (quadratic) polynomials

Definition 1 *Let a function $f(x) : R^n \to R$ and a set of points $Y = \{y^j\}_{j=1}^p \subset R^n$ be given. The (quadratic) polynomial $Q(x)$ is an interpolation of the function $f(x)$ w.r.t. the set Y if*

$$Q(y^j) = f(y^j) \quad j = 1, \ldots, p.$$

Given a linear space of polynomials:

$$\mathcal{L}(\phi_1(x), \cdots, \phi_q(x))$$

Let $Q(y)$ be in this linear space, i.e.,

$$Q(x) = \sum_{i=1}^{q} \alpha_i \phi_i(x),$$

then $Q(x)$ is an interpolation of the function $f(x)$ iff

$$Q(y^j) = \sum_{i=1}^{q} \alpha_i \phi_i(y^j) = f(y^j) \quad j = 1, \ldots, p.$$

This system has a unique solution if $p = q$ and the coefficient matrix is nonsingular (has full-rank).

Interpolation of functions

well poised set of points

The linear system

$$Q(y^j) = \sum_{i=1}^{q} \alpha_i \phi_i(y^j) = f(y^j) \quad j = 1, \ldots, p$$

has a unique solution iff the coefficient matrix

$$\Phi(Y) = \begin{pmatrix} \phi_1(y^1) & \cdots & \phi_q(y^1) \\ \vdots & & \vdots \\ \phi_1(y^p) & \cdots & \phi_q(y^p) \end{pmatrix}$$

of the system is square and nonsingular.

Let $a = (\alpha_1, \ldots, \alpha_q)^T$ and $f(Y) = (f(y^1), \ldots, f(y^p))^T$, then the equation system can be given as

$$\Phi(Y)a = f(Y).$$

Definition 2 *A set of points Y is called* **poised**, *with respect to a given subspace of polynomials, if it can be interpolated by polynomials from this subspace, i.e., if det$(\Phi(Y)) \neq 0$.*

A set of points Y is called **well-poised**, *if it remains poised under small perturbations, i.e., if det$(\Phi(Y))$ is sufficiently large and the matrix $\Phi(Y)$ is well conditioned.*

Quadratic interpolation of functions

the monomial basis

The quadratic monomials

$$\{1, x_1, x_2, \ldots, x_n, x_1^2, \ldots, x_n^2, x_1x_2, \ldots, x_{n-1}x_n\}$$

give a basis in the space of quadratic functions.

The cardinality of the monomial basis is

$$1 + n + n + \frac{n(n-1)}{2} = \frac{(n+1)(n+2)}{2}.$$

Thus we need $\frac{(n+1)(n+2)}{2}$ well-poised points for a full quadratic approximation.

Observe: We can use linear algebra, pivoting, elimination to transform the matrix $\Phi(Y)$ to special form, that means to change the system of basis polynomials.

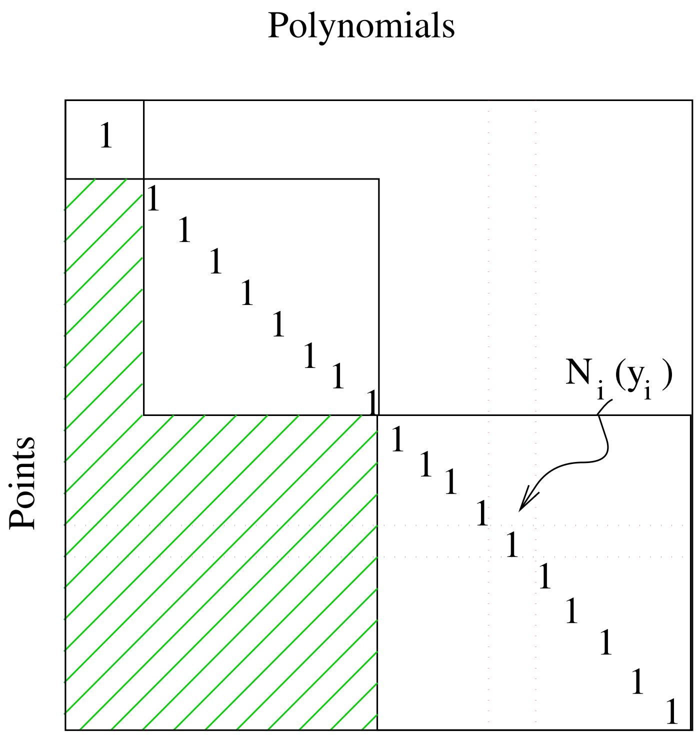
The Lagrange interpolation polynomial

If $\Phi(Y) = I$ then $\phi(y^j) = \ell(y^j)\ \forall\, j$ are the Lagrange interpolation polynomials and $Q(x)$ is the Lagrange interpolation polynomial of $f(x)$.

Quadratic interpolation of functions

the Newton Fundamental polynomials

The basis of quadratic functions and the point set Y is given such a way that $\Phi(Y)$ has the following structure:



Newton fundamental polynomial

Nonzero coefficients $\phi_i(x^j) \neq 0$ occur iff i and j belong to the same subset:

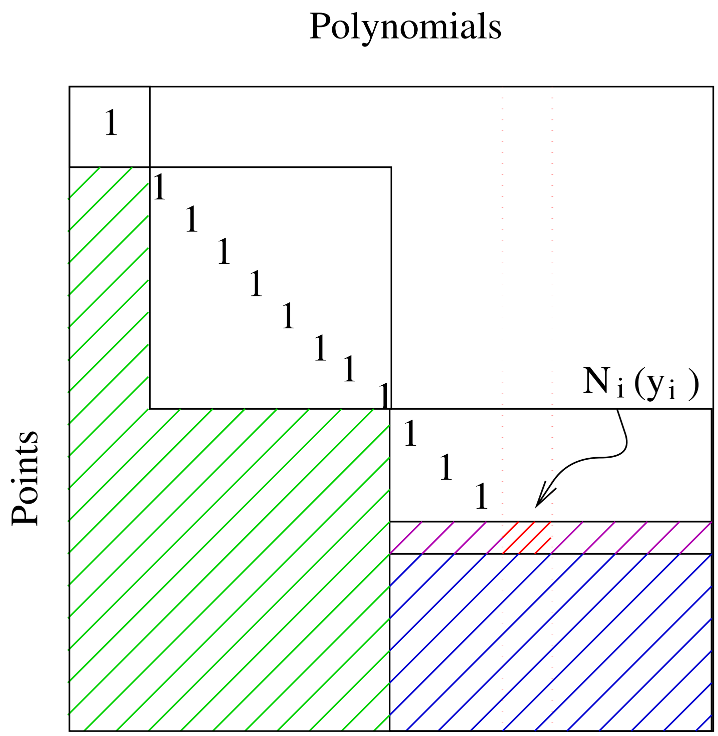
$\{1\}$

$\{x_1, \ldots, x_n\}$

$\{x_1^2, \ldots, x_n^2, x_1 x_2, \ldots, x_{n-1} x_n\}$

Quadratic interpolation of functions

pivoting to find Newton Fundamental polynomials



Pivoting procedure

Initialize $\{\bar{N}_i(\cdot)\}_{i=1}^{p}$.

For $i = 1, \ldots, p$:

> **Normalize:** $\bar{N}_i(x) \leftarrow \frac{\bar{N}_i(x)}{\bar{N}_i(y^i)}$
>
> **Orthogonalize:** $\bar{N}_j(x) \leftarrow \bar{N}_j(x) - \bar{N}_j(y^i)\bar{N}_i(x)$
> for j in the same or "later" block as i

Set $N_i(\cdot) = \bar{N}_i(\cdot)$, $i = 1, \ldots, p$.

The outline of DFO

Quadratic-interpolation Trust-region method

$$\min\ f(x)$$

1. Build a quadratic interpolation $Q(x)$ of $f(x)$. It might be a full quadratic approximation or a partial, but at least full linear.
2. Choose a "trust region radius" Δ and minimize the model function $Q(x)$ around the so-far best point y^1 on this trust region.

$$\widehat{x} := \min \{Q(x) : \|x - y^1\| \leq \Delta\}.$$

3. Update the model.
   - If $\widehat{x}$ is really good, add it to the set, drop another point if needed and increase the radius.
   - If $\widehat{x}$ is good, add it to the set and drop another point if needed.
   - If $\widehat{x}$ is bad, reduce the radius; add the point to the set, if it significantly improves poisedness and drop a point if needed; else reject this point.

We either improve the best solution, the goodness of the best function value, or the goodness of the model.
We can start to build the basis by using as few as two points.

The DFO algorithm–I

Step 0: Initializations.

Let a starting point x^s and the value $f(x^s)$ be given.

Choose an initial trust region radius $\Delta_0 > 0$.

Choose at least one additional point not further than Δ_0 away from x^s to create an initial well-poised interpolation set Y and initial basis of Newton fundamental polynomials.

Determine $x^0 \in Y$ which has the best objective function value; i.e.,

$$f(x^0) = \min_{y^i \in Y} f(y^i).$$

Set $k = 0$.

Set parameters η_0, η_1 to measure progress:

$$0 < \eta_0 < \eta_1 < 1.$$

Step 1: Build the model.

Using the interpolation set Y and the basis of NFP, build an interpolation model $Q_k(x)$.

The DFO algorithm–II

Step 2: Minimize the model within the trust region.

Set $\mathcal{B}_k = \{x : ||x - x^k|| \leq \Delta_k\}$. Compute the point $\widehat{x}^k$ such that

$$Q_k(\widehat{x}^k) = \min_{x \in \mathcal{B}_k} Q_k(x).$$

Compute $f(\widehat{x}^k)$ and the ratio

$$\rho_k \equiv \frac{f(x^k) - f(\widehat{x}^k)}{Q_k(x^k) - Q_k(\widehat{x}^k)}.$$

Step 3: Update the interpolation set.

- If $\rho_k \geq \eta_0$, *include* $\widehat{x}^k$ in Y, dropping one of the existing interpolation points if necessary.
- If $\rho_k < \eta_0$, *include* $\widehat{x}^k$ in Y, if it improves the quality of the model.
- If $\rho_k < \eta_0$ and there are less than $n+1$ points in the intersection of Y and $\mathcal{B}_k$, *generate* a new interpolation point in $\mathcal{B}_k$ while preserving/improving well-poisedness.
- Update the basis of the Newton Fundamental polynomials.

The DFO algorithm–III

Step 4: Update the trust-region radius.

- If $\rho_k \geq \eta_1$, increase the trust region radius

$$\Delta_{k+1} \in [\Delta_k, \gamma_2 \Delta_k].$$

- If $\rho_k < \eta_0$ and cardinality of $Y \cap \mathcal{B}_k$ was less that $n+1$ when $\hat{x}^k$ was computed, reduce the trust region

$$\Delta_{k+1} \in [\gamma_0 \Delta_k, \gamma_1 \Delta_k].$$

- Otherwise, set $\Delta_{k+1} := \Delta_k$.

Step 5: Update the current iterate.

Determine $\bar{x}^k$ with the best objective function value

$$f(\bar{x}^k) = \min_{\substack{y^i \in Y \\ y^i \neq x^k}} f(y^i).$$

If improvement is sufficient (w.r.t. predicted improvement)

$$\bar{\rho}_k \equiv \frac{f(x^k) - f(\bar{x}^k)}{Q_k(x^k) - Q_k(\hat{x}^k)} \geq \eta_0,$$

set $x^{k+1} = \bar{x}^k$.
Else, set $x^{k+1} := x^k$; $k := k+1$; Go to Step 1.

End of algorithm

Goodness of the point set

Improve well-poisedness

Let the set $Y = \{y^1, \cdots, y^{p-1}, y^p\}$ be given.
We want to replace y^p to improve the poisedness of the point set.
This problem results in

$$
\begin{array}{ll}
\max & \det(\Phi(y)) = \begin{pmatrix} \phi_1(y^1) & \cdots & \phi_p(y^1) \\ \vdots & & \vdots \\ \phi_1(y^{p-1}) & \cdots & \phi_p(y^{p-1}) \\ \phi_1(y) & \cdots & \phi_p(y) \end{pmatrix} \\
\text{s.t.} & \|y - y^1\| \le \Delta.
\end{array}
$$

The function

$$
\det(\Phi(y)) = \begin{pmatrix} \phi_1(y^1) & \cdots & \phi_p(y^1) \\ \vdots & & \vdots \\ \phi_1(y^{p-1}) & \cdots & \phi_p(y^{p-1}) \\ \phi_1(y) & \cdots & \phi_p(y) \end{pmatrix} = \ell_p(y)
$$

is the Lagrange function, hence independent of the basis!
Let y^* be the maximizer of this problem.

Let $\bar{Y} = \{y^1, \cdots, y^{p-1}, y^*\}$.

Accept y^* **if** $\Phi(\bar{Y}) \ge 2\Phi(Y)$
or if adding y^* increases the cardinality of point set and $\Phi(\bar{Y})$ is sufficiently large.