Supplementary information
Convergent algorithms for protein structural alignment

Leandro Martínez, José Mario Martínez, Roberto Andreani
State University of Campinas

Here we give a detailed description of the Newtonian algorithm used in DP-Newton and NB-Newton. Although the original problem is given in terms of maximization the maximum of a set of functions, here we use the "minimization of the minimum" approach, which is trivially equivalent.

Problem

$$\text{Minimize } f_{min}(x)$$

where

$$f_{min}(x) = \min\{f_1(x), \ldots, f_m(x)\}.$$

We denote $I_{min}(x) = \{i \in \{1, \ldots, m\} \mid f_i(x) = f_{min}(x)\}$.

Algorithm U1. Let $\theta \in (0,1), \alpha \in (0,1), M > 1, \beta > 0, t_{one} > 0$ be algorithmic parameters. Let $x_0 \in I\!R^n$ be the initial approximation. Given $x_k \in I\!R^n$, the steps for computing x_{k+1} are:

Step 1. Choose $\nu(k) \in I_{min}(x_k)$. If $\|\nabla f_{\nu(k)}(x_k)\| = 0$, terminate.
Step 2. Compute $d_k \in I\!R^n$ such that

$$\nabla f_{\nu(k)}(x_k)^T d_k \leq -\theta \|d_k\| \|\nabla f_{\nu(k)}(x_k)\| \tag{1}$$

and

$$\|d_k\| \geq \beta \|\nabla f_{\nu(k)}(x_k)\|. \tag{2}$$

In the Newton version of the algorithm we choose

$$d = -(\nabla^2 f_{\nu(k)}(x_k) + \lambda I)^{-1} \nabla f_{\nu(k)}(x_k), \tag{3}$$

where λ is the first number in the sequence $\{0, 0.1\|\nabla^2 f_{\nu(k)}(x_k)\|, 0.2\|\nabla^2 f_{\nu(k)}(x_k)\|, \ldots\}$ that verifies (1).

The choice (3) corresponds to take $d = \hat{x} - x_k$ where $\hat{x}$ is the minimizer of a quadratic approximation $q(x)$ of $f_{\nu(k)}(x)$. Namely,

$$q(x) = f(x_k) + \nabla f_{\nu(k)}(x_k)^T (x - x_k) + \frac{1}{2}(x - x_k)^T (\nabla^2 f_{\nu(k)}(x_k) + \lambda I)(x - x_k).$$

If $\lambda = 0$ (which is the usual case) this is the ordinary Taylor approximation of $f_{\nu(k)}$. Sometimes it is necessary to take $\lambda > 0$ in order to guarantee that a minimum of the quadratic exists and that the generated direction is a descent direction (1). In this case, the geometrical meaning of d is that $d = \hat{x} - x_k$, where $\hat{x}$ minimizes the Taylor quadratic approximation in a restricted trust ball [2].

If d satisfies (2) we take $d_k = d$. Otherwise, we take $d_k = \beta d_k / \|d_k\|$.

Step 3. Compute $t_k > 0$, $x_{k+1} \in I\!R^n$, such that

$$f_{min}(x_{k+1}) \leq f_{min}(x_k) + \alpha t_k \nabla f_{\nu(k)}(x_k)^T d_k \tag{4}$$

and

$$\Big[t_k \geq t_{one}\Big] \quad \text{or} \quad \Big[f_{min}(x_k + \bar{t}_k d_k) > f_{min}(x_k) + \alpha \bar{t}_k \nabla f_{\nu(k)}(x_k)^T d_k \ \text{ for some } \bar{t}_k \leq M t_k\Big]. \tag{5}$$

In the Newton implementation of **Step 3**, we use $t_{one} = 1$, $\theta = 10^{-4}$, $\beta = 10^{-6}$ and we proceed according to:

1. $t \leftarrow 1$;

2. **Sufficient Descent Test**. If t satisfies

$$f_{min}(x_{k+1}) \leq f_{min}(x_k) + \alpha t \nabla f_{\nu(k)}(x_k)^T d_k \tag{6}$$

   take $t_k = t$, $x_{k+1} = x_k + t_k d_k$ and finish Step 3.

3. If t does not satisfy (6) compute $\hat{t}$ as:

$$\hat{t} = \frac{-\nabla f_{\nu(k)}(x_k)^T d_k \; t^2}{2(f_{min}(x_k + t d_k) - f_{min}(x_k) - \nabla f_{\nu(k)}(x_k)^T d_k \; t)}. \tag{7}$$

   (If the denominator of the expression above vanishes, we take $\hat{t} = 0.5$.)

   With the choice (7), $\hat{t}$ is the minimizer of the one-dimensional quadratic (parabola) φ that interpolates f_{min} at x_k and $x_k + t d_k$ along the direction d_k. By this we mean that

$$\varphi(0) = f_{min}(x_k), \varphi'(0) = \nabla f_{\nu(k)}(x_k)^T d_k, \varphi(t) = f_{min}(x_k + t d_k).$$

   If $\hat{t} > t/2$ we take $t \leftarrow t/2$. If $\hat{t} < t/10$ we take $t \leftarrow t/10$. Otherwise, take $t \leftarrow \hat{t}$. (This procedure is known as *safeguarded quadratic interpolation* [3].) Go to **Sufficient Descent Test**.

We say that x_* is a critical point if $\nabla f_i(x) = 0$ for some $i \in I_{min}(x)$. Critical points are Clarke Stationary points in the sense used in [1], for example.

In the following theorems we prove that the algorithm stops at x_k only if x_k is critical and that limit points of sequences generated by Algorithm **U1** are critical.

Theorem 1. *Algorithm* **U1** *is well-defined and terminates at x_k only if x_k is critical.*

Proof. Assume that x_k is not critical and define $i = \nu(k)$. So, $\nabla f_i(x_k) \neq 0$. By (1) and the differentiability of f_i,

$$\lim_{t \to 0} \frac{f_i(x_k + t d_k) - f_i(x_k)}{t} = \nabla f_i(x_k)^T d_k < 0.$$

Then,

$$\lim_{t\to 0} \frac{f_i(x_k + td_k) - f_i(x_k)}{t\nabla f_i(x_k)^T d_k} = 1.$$

Since $\alpha < 1$, for t small enough we have:

$$\frac{f_i(x_k + td_k) - f_i(x_k)}{t\nabla f_i(x_k)^T d_k} \geq \alpha.$$

Since $\nabla f_i(x_k)^T d_k < 0$, we deduce:

$$f_i(x_k + td_k) \leq f_i(x_k) + \alpha t \nabla f_i(x_k)^T d_k.$$

But $f_{min}(x_k + td_k) \leq f_i(x_k + td_k)$ and $f_{min}(x_k) = f_i(x_k)$, so:

$$f_{min}(x_k + td_k) \leq f_{min}(x_k) + \alpha t \nabla f_i(x_k)^T d_k \tag{8}$$

for t small enough.

Therefore, choosing t_k small enough, the conditions (4) and (5) are satisfied.

This proves that, whenever x_k is not critical, a point x_{k+1} satisfying (4)-(5) may be found, so the algorithm is well defined. □

Theorem 2 *If x_* is a limit point of a sequence generated by Algorithm* **U1** *then x_* is critical. Moreover, if $\lim_{k\in K} x_k = x_*$ and the same $i = \nu(k) \in I_{min}(x_k)$ is chosen at Step 1 of the algorithm for infinitely many indices $k \in K$, then $i \in I_{min}(x_*)$ and $\nabla f_i(x_*) = 0$. Finally,*

$$\lim_{k\in K} \|\nabla f_{\nu(k)}(x_k)\| = 0. \tag{9}$$

Proof. Let $x_* \in I\!\!R^n$ be a limit point of the sequence generated by Algorithm **U1**. Let $K = \{k_0, k_1, k_2, k_3, \ldots\}$ be an infinite sequence of integers such that:

1. There exists $i \in \{1, \ldots, m\}$ such that $i = \nu(k)$ for all $k \in K$.

2. $\lim_{k\in K} x_k = x_*$.

The sequence K and the index i necessarily exist since $\{1, \ldots, m\}$ is finite.

By the continuity of f_i,

$$\lim_{k\in K} f_i(x_k) = f_i(x_*). \tag{10}$$

Clearly, since $i = \nu(k)$, we have that

$$f_i(x_k) \leq f_\ell(x_k) \ \text{ for all } \ \ell \in \{1, \ldots, m\}.$$

for all $k \in K$.

Taking limits on both sides of this inequality, we see that $f_i(x_*) \leq f_\ell(x_*)$ for all $\ell \in \{1, \ldots, m\}$. Thus,

$$i \in I_{min}(x_*). \tag{11}$$

By the definition of Algorithm **U1**, since $k_{j+1} \geq k_j + 1$, we have:

$$f_i(x_{k_{j+1}})$$

$$= f_{min}(x_{k_{j+1}}) \leq f_{min}(x_{k_j+1}) \leq f_{min}(x_{k_j}) + \alpha t_{k_j} \nabla f_i(x_{k_j})^T d_{k_j} < f_{min}(x_{k_j}) = f_i(x_{k_j}) \tag{12}$$

for all $j \in I\!N$.

By (4), (10) and (12), we obtain:

$$\lim_{j \to \infty} t_{k_j} \nabla f_i(x_{k_j})^T d_{k_j} = 0.$$

Therefore, by (1),

$$\lim_{j \to \infty} t_{k_j} \|\nabla f_i(x_{k_j})\| \|d_{k_j}\| = 0. \tag{13}$$

If, for some subsequence $K_1 \subset K$, $\lim_{k \in K_1} \nabla f_i(x_k) = 0$, we deduce that $\nabla f_i(x_*) = 0$ and the thesis is proved. Therefore, we only need to analyze the possibility that $\|\nabla f_i(x_k)\|$ is bounded away from zero for $k \in K$. In this case, by (13),

$$\lim_{k \in K} t_k \|d_k\| = 0. \tag{14}$$

If, for some subsequence, $\|d_k\| \to 0$, the condition (1) also implies that $\nabla f_i(x_k) \to 0$ and $\nabla f_i(x_*) = 0$. Thus, we only need to consider the case in which $\lim_{k \in K} t_k = 0$. Without loss of generality, we may assume that $t_k < t_{one}$ for all $k \in K$. So, by (5), for all $k \in K$ there exists $\bar{t}_k > 0$ such that

$$f_i(x_k + \bar{t}_k d_k) \geq f_{min}(x_k + \bar{t}_k d_k) > f_{min}(x_k) + \alpha \bar{t}_k \nabla f_i(x_k)^T d_k = f_i(x_k) + \alpha \bar{t}_k \nabla f_i(x_k)^T d_k. \tag{15}$$

Moreover, by (5) and (14),

$$\lim_{k \in K} \bar{t}_k \|d_k\| = 0. \tag{16}$$

Define $s_k = \bar{t}_k d_k$ for all $k \in K$. Then, by (16),

$$\lim_{k \in K} \|s_k\| = 0. \tag{17}$$

By (15) and the Mean Value Theorem, for all $k \in K$ there exists $\xi_k \in [0, 1]$ such that

$$\nabla f_i(x_k + \xi_k s_k)^T s_k = f_i(x_k + s_k) - f_i(x_k) > \alpha \nabla f_i(x_k)^T s_k. \tag{18}$$

Moreover, by (1),

$$\frac{\nabla f_i(x_k)^T s_k}{\|s_k\|} \leq -\theta \|\nabla f_i(x_k)\| \tag{19}$$

for all $k \in K$.

Let $K_1 \subset K$, $s \in I\!R^n$ be such that $\lim_{k \in K_1} s_k / \|s_k\| = s$.

By (17), dividing both sides of the inequality (18) by $\|s_k\|$, and taking limits for $k \in K_1$, we obtain:

$$\nabla f_i(x_*)^T s \geq \alpha \nabla f_i(x_*)^T s.$$

Since $\alpha < 1$ and $\nabla f_i(x_k)^T d_k < 0$ for all k, this implies that $\nabla f_i(x_*)^T s = 0$. Thus, taking limits in (19), we obtain that $\nabla f_i(x_*) = 0$. Therefore, by (11), x_* is critical.

Finally, let us prove (9). If (9) is not true, there exists j and an infinite set of indices $k \in K$ such that $j = \nu(k)$ and $\|\nabla f_j(x_k)\|$ is bounded away from zero. This implies that $j \in I_{min}(x_*)$ and $\|\nabla f_j(x_*)\| \neq 0$, contradicting the first part of the proof. $\square$

References

[1] Audet, C., Dennis, J. E. (2006) Mesh adaptive direct search algorithms for constrained optimization. SIAM J. Optimiz., **17**, 188-217.

[2] A. R. Conn, N. I. M. Gould and Ph. L. Toint, *Trust-Region Methods*, MPS/SIAM Series on Optimization, SIAM-MPS, Philadelphia, 2000.

[3] Dennis, J. E., Schnabel, R. B. (1983) *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice Hall, Englewood Cliffs, NJ.