Sobolev seminorm of quadratic functions with applications to derivative-free optimization

Zaikun Zhang

Abstract This paper studies the H^1 Sobolev seminorm of quadratic functions. The research is motivated by the least-norm interpolation that is widely used in derivative-free optimization. We express the H^1 seminorm of a quadratic function explicitly in terms of the Hessian and the gradient when the underlying domain is a ball. The seminorm gives new insights into least-norm interpolation. It clarifies the analytical and geometrical meaning of the objective function in least-norm interpolation. We employ the seminorm to study the extended symmetric Broyden update proposed by Powell. Numerical results show that the new thoery helps improve the performance of the update. Apart from the theoretical results, we propose a new method of comparing derivative-free solvers, which is more convincing than merely counting the numbers of function evaluations.

Keywords Sobolev seminorm · Least-norm interpolation · Derivative-free optimization · Extended symmetric Broyden update

Mathematics Subject Classification (2010) 90C56 · 90C30 · 65K05

1 Motivation and Introduction

Consider an unconstrained derivative-free optimization problem

$$\min_{x\in\mathbb{R}^n} F(x), \tag{1.1}$$

where F is a real-valued function whose derivatives are unavailable. Problems of this type have numerous applications. For instance, they have been employed to solve the helicopter rotor blade design problem [5,4,3], the ground water community problems [23,20], and the problems in biomedical imaging [30,31].

Zaikun Zhang (www.zhangzk.net)
Institute of Computational Mathematics and Scientific/Engineering Computing, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, P.O. Box 2719, Beijing 100190, CHINA
E-mail: zhangzk@lsec.cc.ac.cn / zaikunzhang@gmail.com

Many algorithms have been developed for problem (1.1). For example, direct search methods (see Wright [47], Powell [34], Lewis, Torczon, and Trosset [28], and Kolda, Lewis, and Torczon [27] for reviews), line search methods without derivatives (see Stewart [42] for a quasi-Newton method using finite difference; see Gilmore and Kelley [21], Choi and Kelley [7], and Kelley [25,26] for Implicit Filtering method, a hybrid of quasi-Newton and grid-based search; see Diniz-Ehrhardt, Martínez, and Raydán [16] for a derivative-free line search technique), and model-based methods (for instance, a method by Winfield [46], COBYLA by Powell [33], DFO by Conn, Scheinberg, and Toint [9,10,11], UOBYQA by Powell [35], a wedge trust region method by Marazzi and Nocedal [29], NEWUOA and extensions by Powell [38,39,40,41], CONDOR by Vanden Berghen and Bersini [43], BOOSTERS by Oeuvray and Bierlaire [32], ORBIT by Wild, Regis, and Shoemaker [45], MNH by Wild [44], and a recent algorithm using sparse low degree model by Bandeira, Scheinberg, and Vicente [2]). We refer the readers to the book by Brent [6], the one by Kelley [24], and the one by Conn, Scheinberg, and Vicente [12] for extensive discussions and the references therein.

Multivariate interpolation has been acting as a powerful tool in the design of derivative-free optimization methods. The following quadratic interpolation plays an important role in Conn and Toint [13], Conn, Scheinberg, and Toint [9,10,11], Powell [37,38,39,40,41], and Custódio, Rocha, and Vicente [14]:

$$\min_{Q \in \mathcal{Q}} \|\nabla^2 Q\|_{\mathrm{F}}^2 + \sigma \|\nabla Q(x_0)\|_2^2$$
$$\text{s.t. } Q(x) = f(x), \quad x \in \mathcal{S}, \tag{1.2}$$

where $\mathcal{Q}$ is the linear space of polynomials with degree at most two in n variables, x_0 is a specific point in $\mathbb{R}^n$, σ is a nonnegative constant, $\mathcal{S}$ is a finite set of interpolation points in $\mathbb{R}^n$, and f is a function[1] on $\mathcal{S}$. We henceforth refer to this type of interpolation as least-norm interpolation.

The objective function of problem (1.2) is interesting. It is easy to handle, and practice has proved it successful. The main purpose of this paper is to provide a new interpretation of it. Our tool is the H^1 Sobolev seminorm, which is classical in PDE theory but may be rarely noticed in nonlinear optimization. We will give new insights into some basic questions about the objective function in (1.2). For example, *what is the exact analytical and geometrical meaning of this objective*? Why to combine the Frobenius norm of the Hessian and the 2-norm of the gradient? What is the meaning of the parameters x_0 and σ?

This paper is organized as follows. Section 2 introduces the applications of least-norm interpolation in derivative-free optimization. Section 3 presents the main theoretical results. We first study the H^1 seminorm of quadratic functions, and then employ the H^1 seminorm to investigate least-norm interpolation. The questions listed above are answered in Section 3. Section 4 applies the theory obtained in Section 3 to study the extended symmetric Broyden update proposed by Powell [41]. We show an easy and effective way to improve the performance of the update. Section 5 concludes our discussion.

The main contributions of the paper lie in Section 3 and Section 4. Besides the theoretical results, the numerical experience in Section 4 is also a highlight. We propose a new method of testing derivative-free solvers by introducing statistics into the numerical experiments. See Subsection 4.3 for details.

---

[1] Notice that f is not always equal to the objective function F, which is the case in Powell [37,38,39,40, 41]. See Section 2 for details.

A remark on notation. We use the notation

$$
\begin{aligned}
\min_{x \in X} \ & \psi(x) \\
\text{s.t. } & \min_{x \in X} \phi(x)
\end{aligned}
\tag{1.3}
$$

for the bilevel programming problem

$$
\begin{aligned}
\min \ & \psi(x) \\
\text{s.t. } & x \in [\arg\min_{\xi \in X} \phi(\xi)].
\end{aligned}
\tag{1.4}
$$

2 Least-Norm Interpolation in Derivative-Free Optimization

Least-norm interpolation has successful applications in derivative-free optimization, especially in model-based algorithms. These algorithms typically follow trust-region methodology [8,36]. On each iteration, a local model of the objective function is constructed, and then it is minimized within a trust-region to generate a trial step. The model is usually a quadratic polynomial constructed by solving an interpolation problem

$$
Q(x) = F(x), \quad x \in \mathcal{S},
\tag{2.1}
$$

where, as stated in problem (1.1), F is the objective function, and $\mathcal{S}$ is an interpolation set. To determine a unique quadratic polynomial by problem (2.1), the size of $\mathcal{S}$ needs to be at least $(n+1)(n+2)/2$, which is prohibitive when n is big. Thus we need to consider underdetermined quadratic interpolation. In that case, a classical strategy to take up the remaining freedom is to minimize some functional subject to the interpolation constraints, that is to solve

$$
\begin{aligned}
\min_{Q \in \mathcal{Q}} \ & \mathcal{F}(Q) \\
\text{s.t. } & Q(x) = F(x), \quad x \in \mathcal{S},
\end{aligned}
\tag{2.2}
$$

$\mathcal{F}$ being some functional on $\mathcal{Q}$. Several existing choices of $\mathcal{F}$ lead to least-norm interpolation, directly or indirectly. Here we give some examples.

Conn and Toint [13] suggests the quadratic model that solves

$$
\begin{aligned}
\min_{Q \in \mathcal{Q}} \ & \|\nabla^2 Q\|_{\mathrm{F}}^2 + \|\nabla Q(x_0)\|_2^2 \\
\text{s.t. } & Q(x) = F(x), \quad x \in \mathcal{S},
\end{aligned}
\tag{2.3}
$$

where x_0 is a specific point in $\mathbb{R}^n$ (the center of the trust-region they use). Problem (2.3) is a least-norm interpolation problem with $\sigma = 1$.

Conn, Scheinberg, and Toint [9,10,11] builds a quadratic model by solving

$$
\begin{aligned}
\min_{Q \in \mathcal{Q}} \ & \|\nabla^2 Q\|_{\mathrm{F}}^2 \\
\text{s.t. } & Q(x) = F(x), \quad x \in \mathcal{S}.
\end{aligned}
\tag{2.4}
$$

It is considered as the best general strategy for the DFO algorithm to select a unique model from the pool of the possible interpolation models [11]. Wild [44] also works with the model defined by (2.4). Problem (2.4) is a least-norm interpolation problem with $\sigma = 0$. Conn, Scheinberg, and Vicente [12] explains the motivation for (2.4) by the following error bounds of quadratic interpolants.

Theorem 2.1 [12] *Suppose that the interpolation set $S = \{y_0, y_1, \ldots, y_m\}$ $(m \geq n)$ is contained in a ball $\mathcal{B}(y_0, r)$ $(r > 0)$, and the matrix*

$$L = \frac{1}{r}(y_1 - y_0 \cdots y_m - y_0) \tag{2.5}$$

has rank n. If F is continuously differentiable on $\mathcal{B}(y_0, r)$, and ∇F is Lipschitz continuous on $\mathcal{B}(y_0, r)$ with constant $\nu > 0$, then for any quadratic function Q satisfying the interpolation constraints (2.1), it holds that

$$\|\nabla Q(x) - \nabla F(x)\|_2 \leq \frac{5\sqrt{m}}{2}\|L^\dagger\|_2(\nu + \|\nabla^2 Q\|_2)r, \quad \forall x \in \mathcal{B}(y_0, r), \tag{2.6}$$

and

$$|Q(x) - F(x)| \leq \left(\frac{5\sqrt{m}}{2}\|L^\dagger\|_2 + \frac{1}{2}\right)(\nu + \|\nabla^2 Q\|_2)r^2, \quad \forall x \in \mathcal{B}(y_0, r). \tag{2.7}$$

In light of Theorem 2.1, minimizing some norm of $\nabla^2 Q$ will help improve the approximation of gradient and function value. Notice that $\nabla^2 Q$ appears in (2.6–2.7) with 2-norm rather than Frobenius norm.

Powell [37, 38, 39] introduce the symmetric Broyden update to derivative-free optimization, and it is proposed to solve

$$\begin{aligned} \min_{Q \in \mathcal{Q}} &\ \|\nabla^2 Q - \nabla^2 Q_0\|_F^2 \\ \text{s.t. } &\ Q(x) = F(x), \ \ x \in S \end{aligned} \tag{2.8}$$

to obtain a model for the current iteration, provided that Q_0 is the quadratic model used in the previous iteration. Problem (2.8) is essentially a least-norm interpolation problem about $Q - Q_0$. The symmetric Broyden update is motivated by least change secant updates [15] in quasi-Newton methods. One particularly interesting advantage of the update is that, when F is a quadratic function, $\nabla^2 Q_+$ approximates $\nabla^2 F$ better than $\nabla^2 Q_0$ unless $\nabla^2 Q_+ = \nabla^2 Q_0$, Q_+ being the model obtained by the update [37, 38, 39].

Powell [41] proposes the extended symmetric Broyden update by adding a first-order term to the objective function of problem (2.8), resulting in

$$\begin{aligned} \min_{Q \in \mathcal{Q}} &\ \|\nabla^2 Q - \nabla^2 Q_0\|_F^2 + \sigma\|\nabla Q(x_0) - \nabla Q_0(x_0)\|_2^2 \\ \text{s.t. } &\ Q(x) = F(x), \ \ x \in S, \end{aligned} \tag{2.9}$$

x_0 and σ being specifically selected parameters, σ nonnegative. Again, (2.9) can be interpreted as a least-norm interpolation about $Q - Q_0$. Powell [41] motivates (2.9) by an algebraic example for which the symmetric Broyden update does not behave satisfactorily. We will study the extended symmetric Broyden update in Section 4.

Apart from model-based methods, least-norm interpolation also has applications in direct search methods. Custódio, Rocha, and Vicente [14] incorporates models defined by (2.4) and (2.8) into direct search. The authors attempt to enhance the performance of direct search methods by taking search steps based on these models. It is reported that (2.4) works better for their method, and their procedure provides significant improvements to direct search methods of directional type.

For more discussions about least-norm interpolation in derivative-free optimization, we refer the readers to Chapters 5 and 11 of Conn, Scheinberg, and Vicente [12].

3 The H^1 Sobolev Seminorm of Quadratic Functions

In Sobolev space theory [1, 18], the H^1 seminorm of a function f over a domain Ω is defined as

$$|f|_{H^1(\Omega)} = \left[\int_\Omega \|\nabla f(x)\|_2^2 \, dx \right]^{1/2}. \tag{3.1}$$

In this section, we give an explicit formula for the H^1 seminorm of quadratic functions when Ω is a ball, and accordingly present a new understanding of least-norm interpolation. We prove that least-norm interpolation essentially seeks the quadratic interpolant with minimal H^1 seminorm over a ball. Theorem 3.1, Theorem 3.2, and Theorem 3.3 are the main theoretical results of this paper.

3.1 The H^1 Seminorm of Quadratic Functions

The H^1 seminorm of a quadratic function over a ball can be expressed explicitly in terms of its coefficients. We present the formula in the following theorem.

Theorem 3.1 *Let x_0 be a point in $\mathbb{R}^n$, r be a positive number, and*

$$\mathcal{B} = \{x \in \mathbb{R}^n : \|x - x_0\|_2 \leq r\}. \tag{3.2}$$

Then for any $Q \in \mathcal{Q}$,

$$|Q|_{H^1(\mathcal{B})}^2 = V_n r^n \left[\frac{r^2}{n+2} \|\nabla^2 Q\|_F^2 + \|\nabla Q(x_0)\|_2^2 \right], \tag{3.3}$$

where V_n is the volume of the unit ball in $\mathbb{R}^n$.

Proof Without loss of generality, we assume that $x_0 = 0$. Let

$$g = \nabla Q(0), \text{ and } G = \nabla^2 Q. \tag{3.4}$$

Then

$$|Q|_{H^1(\mathcal{B})}^2 = \int_{\|x\|_2 \leq r} \|Gx + g\|_2^2 \, dx = \int_{\|x\|_2 \leq r} \left(x^T G^2 x + 2x^T G g + \|g\|_2^2 \right) dx. \tag{3.5}$$

Because of symmetry, the integral of $x^T G g$ is zero. Besides,

$$\int_{\|x\|_2 \leq r} \|g\|_2^2 \, dx = V_n r^n \|g\|_2^2. \tag{3.6}$$

Thus we only need to find the integral of $x^T G^2 x$. Now we assume that G is diagonal (if not, apply a rotation). Denote the i-th diagonal entry of G by $G_{(ii)}$, and the i-th coordinate of x by $x_{(i)}$. Then

$$\int_{\|x\|_2 \leq r} x^T G^2 x \, dx = \int_{\|x\|_2 \leq r} \left[\sum_{i=1}^n G_{(ii)}^2 x_{(i)}^2 \right] dx = \sum_{i=1}^n \left[G_{(ii)}^2 \int_{\|x\|_2 \leq r} x_{(i)}^2 \, dx \right]. \tag{3.7}$$

To finish the proof, we show that

$$\int_{\|x\|_2 \leq r} x_{(i)}^2 \, dx = \frac{V_n r^{n+2}}{n+2}. \tag{3.8}$$

It suffices to justify (3.8) for the case $r = 1$ and $n \geq 2$. First,

$$
\begin{aligned}
& \int_{\|x\|_2 \leq 1} x_{(i)}^2 \, \mathrm{d}x \\
&= \int_{-1}^{1} u^2 \, \mathrm{d}u \int_{\|v\|_2 \leq \sqrt{1-u^2}} \mathrm{d}v \ \ (v \in \mathbb{R}^{n-1}) \\
&= V_{n-1} \int_{-1}^{1} u^2 (1-u^2)^{\frac{n-1}{2}} \, \mathrm{d}u \\
&= V_{n-1} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sin^2 \theta \cos^n \theta \, \mathrm{d}\theta,
\end{aligned}
\tag{3.9}
$$

and, similarly,

$$
V_n = V_{n-1} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos^n \theta \, \mathrm{d}\theta.
\tag{3.10}
$$

Second, integration by parts shows that

$$
\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \cos^{n+2} \theta \, \mathrm{d}\theta = (n+1) \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \sin^2 \theta \cos^n \theta \, \mathrm{d}\theta.
\tag{3.11}
$$

Now it is easy to obtain (3.8) from (3.9-3.11). $\qquad\square$

Theorem 3.1 tells us that the H^1 seminorm of a quadratic function Q over a ball $\mathcal{B}$ is closely related to a combination of $\|\nabla^2 Q\|_F^2$ and $\|\nabla Q(x_0)\|_2^2$, x_0 being the center of $\mathcal{B}$, and the combination coefficients being determined by the radius of $\mathcal{B}$. This result is interesting for two reasons. First, it enables us to measure the overall magnitude of the gradient over a ball. This is nontrivial for a quadratic function, since its gradient is not constant. Second, it clarifies the analytical and geometrical meaning of combining the Frobenius norm of the Hessian and the 2-norm of the gradient, and enables us to select the combination coefficients according to the geometrical meaning.

3.2 New Insights into Least-Norm Interpolation

The H^1 seminorm provides a new angle of view to understand least-norm interpolation. For convenience, we rewrite the least-norm interpolation problem here and call it the problem $P_1(\sigma)$:

$$
\begin{aligned}
& \min_{Q \in \mathcal{Q}} \ \|\nabla^2 Q\|_F^2 + \sigma \|\nabla Q(x_0)\|_2^2 \\
& \text{s.t. } Q(x) = f(x), \ \ x \in \mathcal{S}.
\end{aligned}
\tag{$P_1(\sigma)$}
$$

We assume that the interpolation system

$$
Q(x) = f(x), \ \ x \in \mathcal{S}
\tag{3.12}
$$

is consistent on $\mathcal{Q}$, namely

$$
\{Q \in \mathcal{Q} : Q(x) = f(x) \text{ for every } x \in \mathcal{S}\} \neq \emptyset.
\tag{3.13}
$$

With the help of Theorem 3.1, we see that the problem $P_1(\sigma)$ is equivalent to the problem

$$\min_{Q\in\mathcal{Q}} |Q|_{H^1(\mathcal{B}_r)}$$
$$\text{s.t. } Q(x) = f(x), \quad x \in \mathcal{S} \tag{$P_2(r)$}$$

in some sense. The purpose of this subsection is to clarify the equivalence.

When $\sigma > 0$, the equivalence is easy and we state it as follows.

Theorem 3.2 *If $\sigma > 0$, then the least-norm interpolation problem $P_1(\sigma)$ is equivalent to the problem $P_2(r)$ with $r = \sqrt{(n+2)/\sigma}$.*

It turns out that the least-norm interpolation problem with positive σ essentially seeks the interpolant with minimal H^1 seminorm over a ball. The geometrical meaning of x_0 and σ is clear now: x_0 is the center of the ball, and $\sqrt{(n+2)/\sigma}$ is the radius.

Now we consider the least-norm interpolation problem with $\sigma = 0$. This case is particularly interesting, because it appears in several practical algorithms [11,38,44]. Since $P_1(\sigma)$ may have multiple solutions when $\sigma = 0$, we modify the definition of $P_1(0)$: from now on, $P_1(0)$ is defined to be the bilevel least-norm interpolation problem

$$\min_{Q\in\mathcal{Q}} \|\nabla Q(x_0)\|_2$$
$$\text{s.t. } \min_{Q\in\mathcal{Q}} \|\nabla^2 Q\|_F \tag{$P_1(0)$}$$
$$\text{s.t. } Q(x) = f(x), \quad x \in \mathcal{S}.$$

The new definition is reasonable, because the following proposition holds after the modification.

Proposition 3.1 *For each $\sigma \geq 0$, the problem $P_1(\sigma)$ has a unique solution Q_σ. Moreover, Q_σ converges[2] to Q_0 when σ tends to 0^+.*

Proof The uniqueness of the solution is simple. First, the uniqueness of the Hessian and gradient is guaranteed by the convexity of the objective functional(s); then with the help of any one of the interpolation constraints, we find that the constant term is also unique.

In light of the uniqueness stated above, the convergence of $\{Q_\sigma\}$ is a corollary of the classical penalty function theory in nonlinear optimization (see, for instance, Theorem 12.1.1 of Fletcher [19]). □

Note that Proposition 3.1 implies the problem $P_2(r)$ has a unique solution for each positive r. Now we can state the relation between the problems $P_1(0)$ and $P_2(r)$.

Theorem 3.3 *When r tends to infinity, the solution of $P_2(r)$ converges to the solution of $P_1(0)$.*

Theorem 3.3 indicates that, when $\sigma = 0$, the least-norm interpolation problem seeks the interpolant with minimal H^1 seminorm over $\mathbb{R}^n$ in the sense of limit. Thus Theorem 3.3 is the extension of Theorem 3.2 to the case $\sigma = 0$.

The questions listed in Section 1 can be answered now. The meaning of the objective function in least-norm interpolation is to minimize the H^1 seminorm of the interpolant over a ball. The reason to combine the Frobenius norm of the Hessian and the 2-norm of the gradient is to measure the H^1 seminorm of the interpolant. The parameters x_0 and σ determines the ball where the H^1 seminorm is calculated, x_0 being the center and $\sqrt{(n+2)/\sigma}$ being the radius. When $\sigma = 0$, we can interprete least-norm interpolation in the sence of limit.

---

[2] We define the convergence on $\mathcal{Q}$ to be the convergence of coefficients.

4 On the Extended Symmetric Broyden Update

In this section, we employ the H^1 seminorm to study the extended symmetric Broyden update proposed by Powell [41]. As introduced in Section 2, the update defines Q_+ to be the solution of

$$\min_{Q \in \mathcal{Q}} \|\nabla^2 Q - \nabla^2 Q_0\|_F^2 + \sigma \|\nabla Q(x_0) - \nabla Q_0(x_0)\|_2^2$$
$$\text{s.t. } Q(x) = F(x), \ x \in \mathcal{S}, \tag{4.1}$$

F being the objective function, and Q_0 being the model used in the previous trust-region iteration. When $\sigma = 0$, it is the symmetric Broyden update studied by Powell [37,38,39]. We focus on the case $\sigma > 0$. In Subsection 4.1, we interpret the update with the H^1 seminorm. In Subsection 4.2, we discuss the choices of x_0 and σ in the update, and show how to choose x_0 and σ according to their geometrical meaning. In Subsection 4.3, we test our choices of x_0 and σ through numerical experiments. It is worth mentioning that, the numerical experiments in Subsection 4.3 are designed in a special way in order to reduce the influence of computer rounding errors and obtain more convincing results.

4.1 Interpret the Update with the H^1 Seminorm

According to Theorem 3.2, problem (4.1) is equivalent to

$$\min_{Q \in \mathcal{Q}} |Q - Q_0|_{H^1(\mathcal{B})}$$
$$\text{s.t. } Q(x) = F(x), \ x \in \mathcal{S}, \tag{4.2}$$

where

$$\mathcal{B} = \left\{ x \in \mathbb{R}^n : \|x - x_0\|_2 \le \sqrt{(n+2)/\sigma} \right\}. \tag{4.3}$$

Thus the extended symmetric Broyden update seeks the closest quadratic model to Q_0 subject to the interpolation constraints, distance being measured by the seminorm $|\cdot|_{H^1(\mathcal{B})}$.

Similar to the symmetric Broyden update, the extended symmetric Broyden update enjoys a very good approximation property when F is a quadratic function, as stated in Proposition 4.1.

Proposition 4.1 *If F is a quadratic function, then the solution Q_+ of problem (4.1) satisfies*

$$|Q_+ - F|_{H^1(\mathcal{B})}^2 = |Q_0 - F|_{H^1(\mathcal{B})}^2 - |Q_+ - Q_0|_{H^1(\mathcal{B})}^2, \tag{4.4}$$

where $\mathcal{B}$ is defined as (4.3).

Proof Let $Q_t = Q_+ + t(Q_+ - F)$, t being any real number. Then Q_t is a quadratic function interpolating F on $\mathcal{S}$, as F is quadratic and Q_+ interpolates it. The optimality of Q_+ implies that the function

$$\varphi(t) = |Q_t - Q_0|_{H^1(\mathcal{B})}^2 \tag{4.5}$$

attains its minimum when t is zero. Expanding $\varphi(t)$, we obtain

$$\varphi(t) = t^2 |Q_+ - F|_{H^1(\mathcal{B})}^2 + 2t \int_{\mathcal{B}} [\nabla(Q_+ - Q_0)(x)]^{\mathrm{T}} [\nabla(F - Q_0)(x)] \, dx + |Q_+ - Q_0|_{H^1(\mathcal{B})}^2. \tag{4.6}$$

Hence

$$\int_{\mathcal{B}} [\nabla(Q_+ - Q_0)(x)]^{\mathrm{T}} [\nabla(F - Q_0)(x)] \, \mathrm{d}x = 0. \tag{4.7}$$

Then we obtain (4.4) by considering $\varphi(-1)$. □

In light of Theorem 3.1, equation (4.4) is equivalent to equation (1.9) of Powell [41]. Notice that equation (4.4) implies

$$\int_{\mathcal{B}} \|\nabla Q_+(x) - \nabla F(x)\|_2^2 \, \mathrm{d}x \le \int_{\mathcal{B}} \|\nabla Q_0(x) - \nabla F(x)\|_2^2 \, \mathrm{d}x. \tag{4.8}$$

In other words, ∇Q_+ approximates ∇F on $\mathcal{B}$ better than ∇Q_0 unless $\nabla Q_+ = \nabla Q_0$.

4.2 Choices of x_0 and σ

Now we turn our attention to the choices of x_0 and σ in the update. Recall that the purpose of the update is to construct a model for a trust-region subproblem. As in classical trust-region methods, the trust region is available before the update is applied, and we suppose it is

$$\{x : \|x - \bar{x}\|_2 \le \Delta\}, \tag{4.9}$$

the point $\bar{x}$ being the trust-region center, and the positive number Δ being the trust-region radius.

Powell [41] chooses x_0 and σ by exploiting the Lagrange functions of the interpolation problem (4.1). Suppose that $\mathcal{S} = \{y_0, y_1, \ldots, y_m\}$, then the i-th ($i = 0, 1, \ldots, m$) Lagrange function of problem (4.1) is defined to be the solution of

$$\min_{l_i \in \mathcal{Q}} \|\nabla^2 l_i\|_{\mathrm{F}}^2 + \sigma \|\nabla l_i\|_2^2$$
$$\text{s.t. } l_i(y_j) = \delta_{ij}, \quad j = 0, 1, \ldots, m, \tag{4.10}$$

δ_{ij} being the Kronecker delta. In the algorithm of Powell [41], $\mathcal{S}$ is maintained in a way so that Q_0 interpolates F on $\mathcal{S}$ except one point, say y_0 without loss of generality. Then the interpolation constraints can be rewritten as

$$Q(y_j) - Q_0(y_j) = [F(y_0) - Q_0(y_0)] \delta_{0j}, \quad j = 0, 1, \ldots, m. \tag{4.11}$$

Therefore the solution of problem (4.1) is

$$Q_+ = Q_0 + [F(y_0) - Q_0(y_0)] l_0. \tag{4.12}$$

Hence l_0 plays a central part in the update. Thus Powell [41] chooses x_0 and σ by examining l_0. Powell shows that, in order to make sure l_0 behaves well, the ratio $\|x_0 - \bar{x}\|_2 / \Delta$ should not be much larger than one, therefore x_0 is set to be $\bar{x}$ throughout the calculation. The choice of σ is a bit complicated. The basic idea is to balance $\|\nabla^2 l_0\|_{\mathrm{F}}^2$ and $\sigma \|\nabla l_0(x_0)\|_2^2$. Thus σ is set to η / ξ, where η and ξ are estimates of the magnitudes of $\|\nabla^2 l_0\|_{\mathrm{F}}^2$ and $\|\nabla l_0(x_0)\|_2^2$, respectively. See Powell [41] for details.

In Subsection 4.1, we have interpreted the extended symmetric Broyden update with the H^1 seminorm. The new interpretation enables us to choose x_0 and σ in a geometrical way. According to problem (4.2), choosing x_0 and σ is equivalent to choosing the ball $\mathcal{B}$. Let us think about the role that $\mathcal{B}$ plays in the update. First, $\mathcal{B}$ is the region where the update tries to preserve information from Q_0, by minimizing the change with respect to the seminorm

$|\cdot|_{H^1(\mathcal{B})}$. Second, $\mathcal{B}$ is the region where the update tends to improve the model, as suggested by the facts (4.4) and (4.8) for quadratic objective function. Thus we should choose $\mathcal{B}$ to be a region where the behavior of the new model is important to us. It may seem adequate to pick $\mathcal{B}$ to be the trust region (4.9). But we prefer to set it bigger, because the new model Q_+ will influence its successors via the least change update, and thereby influence subsequent trust-region iterations. Thus it is myopic to consider only the current trust region, and a more sensible choice is to let $\mathcal{B}$ be the ball $\{x : \|x - \bar{x}\|_2 \leq M\Delta\}$, M being a positive number bigger than one[3]. Moreover, we find in practice that it is helpful to require $\mathcal{B} \supset \mathcal{S}$. Thus our choice of $\mathcal{B}$ is

$$\{x : \|x - \bar{x}\|_2 \leq r\}, \tag{4.13}$$

where

$$r = \max\left\{M\Delta, \max_{x \in \mathcal{S}} \|x - \bar{x}\|_2\right\}. \tag{4.14}$$

Consequently, we set

$$x_0 = \bar{x}, \quad \text{and} \quad \sigma = \frac{n+2}{r^2}. \tag{4.15}$$

Therefore our choice of x_0 coincides with Powell's, but the choice of σ is different.

4.3 Numerical Results

With the help of the H^1 seminorm, we have proposed a new method of choosing σ for the extended symmetric Broyden update. In this subsection we test the new σ through numerical experiments, and make comparison with the one by Powell [41].

Powell [41] implements a trust-region algorithm based on the extended symmetric Broyden update in Fortran 77. We modify this code to get two solvers as follows for comparison.

a.) ESYMBP: σ is chosen according to Powell [41], as described in the second paragraph of Subsection 4.2.
b.) ESYMBS: σ is chosen according to (4.14–4.15). We set $M = 10$ in (4.14).

In the code of Powell [41], the size of the interpolation set $\mathcal{S}$ is kept unchanged throughout the computation, and the user can set it to be any integer between $n+2$ and $(n+1)(n+2)/2$. We chose $2n + 1$ as recommended. The terminate criterion of this code is determined by a parameter named RHOEND. RHOEND acts as the final trust-region radius in the code, and its magnitude usually agrees with the precision of the computational solution. We tested ESYMBPand ESYMBSwith RHOEND $= 10^{-2}, 10^{-4}$, and 10^{-6}, to observe the performance of the solvers under different requirements of solution precision. The code of Powell [41] is capable of solving problems with box constraints. We set the bounds to infinity since we are focusing on unconstrained problems.

We assume that evaluating the objective function is the most expensive part for optimization without derivatives. Thus we might compare the performance of different derivative-free solvers by simply counting the numbers of function evaluations they use until termination, provided that they have found the same minimizer to high accuracy. However, this is inadequate in practice, because, according to our numerical experiences, computer rounding errors could substantially influence the number of function evaluations. Powell [38] presents

---

[3] Since the trust-region radii of the subsequent iterations vary dynamically according to the degree of success, there should be more elaborate ways to define the domain $\mathcal{B}$ adaptively. The definition given here might be the simplest one, and is enough for our experiment.

Table 4.1: Name of Test Problems

ARGLINA	ARGLINB	ARGLINC	ARWHEAD	BDQRTIC
BROYDN3D	BRYBND	CHROSEN	COSINE	CURLY10
CURLY20	CURLY30	DQRTIC	EDENSCH	EG2
ENGVAL1	FLETCBV2	FREUROTH	GENBROWN	INTEGREQ
NONCVXUN	PENALTY1	PENALTY2	POWER	SCHMVETT
SPARSINE	SPARSQUR	SPHRPTS	TOINTGSS	VARDIM

a very interesting example for this phenomenon. In the example, the computational result of a test problem was influenced dramatically by the order of variables. But the code under test (NEWUOA) was mathematically independent of variables' order. Thus the differences in the results are due entirely to computer rounding errors. Please see Section 8 of Powell [38] for details of the example.

We test the solvers ESYMBS and ESYMBP with the following method, in order to observe their performance in a relatively reliable way, and to inspect their stability with respect to computer rounding errors. The basic idea is from the numerical experiments of Powell [41]. Given a test problem $\mathcal{P}$ with objective function F and starting point $\hat{x}$, we randomly generate N permutation matrices P_i ($i = 1, 2, \ldots, N$), and let

$$F_i(x) = F(P_i x), \ \ \hat{x}_i = P_i^{-1} \hat{x}, \ \ i = 1, 2, \ldots, N. \tag{4.16}$$

Then we employ the solvers to minimize F_i starting form $\hat{x}_i$. For solver $\mathcal{S}$, we obtain a vector $\#F = (\#F_1, \#F_2, \ldots, \#F_N)$, $\#F_i$ being the number of function evaluations required by $\mathcal{S}$ for F_i and $\hat{x}_i$. If the code of $\mathcal{S}$ is mathematically independent of variables' order (it is the case for ESYMBS and ESYMBP), then all the entries of $\#F$ are identical in theory. However, it is not the case in practice, and the differences are completely made by computer rounding errors. We compute the mean value and the standard deviation of vector $\#F$:

$$\text{mean}(\#F) = \frac{1}{N} \sum_{i=1}^{N} \#F_i, \tag{4.17}$$

and

$$\text{std}(\#F) = \sqrt{\frac{1}{N} \sum_{i=1}^{N} [\#F_i - \text{mean}(\#F)]^2}. \tag{4.18}$$

When N is reasonably large, $\text{mean}(\#F)$ estimates the average performance of solver $\mathcal{S}$ on problem $\mathcal{P}$, and $\text{std}(\#F)$ reflects the stability of solver $\mathcal{S}$ with respect to computer rounding errors when solving problem $\mathcal{P}$. We may also compute the relative standard deviation of $\#F$, namely,

$$\text{rstd}(\#F) = \frac{\text{std}(\#F)}{\text{mean}(\#F)}, \tag{4.19}$$

to obtain a normalized measure of stability. By comparing these statistics, we can reasonably assess the solvers under consideration. We use $N = 10$ in practice.

With the method addressed above, we tested ESYMBP and ESYMBS on 30 unconstrained test problems with alterable dimension. The problems are from CUTEr [22] and Powell [38], with names listed in Table 4.1.

For each problem, we solved it for dimensions 6, 8, 10, 12, 14, and 16. We set the maximal number of function evaluations to be 1000. The experiments were carried out on a Thinkpad T400 laptop with Linux 2.6.33, and the compiler was gfortran in GCC 4.3.3.

We visualize the results of our experiments by presenting the performance profiles [17] of mean(#F) in Fig. 4.1, 4.2, and 4.3. The same as traditional performance profiles, the higher means the better. The profiles suggest that, in our experiments, ESYMBS performed better than ESYMBP in the sense of number of function evaluations. Moreover, ESYMBS also showed better stability than ESYMBP, with respect to computer rounding errors. To make the comparison, we present the performance profiles of rstd in Fig. 4.4, 4.5, and 4.6. As before, the higher means the better. Thus the profiles support the conclusion that ESYMBS performed stabler than ESYMBP with respect to computer rounding errors.

From the results of our experiments, we conclude that the H^1 seminorm has suggested a better way of parameter selection for extended symmetric Broyden update, compared with the method given in Powell [41]. The new selection reduces the number of function evaluations in the resultant algorithm, and improves the solver's stability with respect to computer rounding errors. Moreover, the new selection is much simpler and easier to understand than the one in Powell [41].
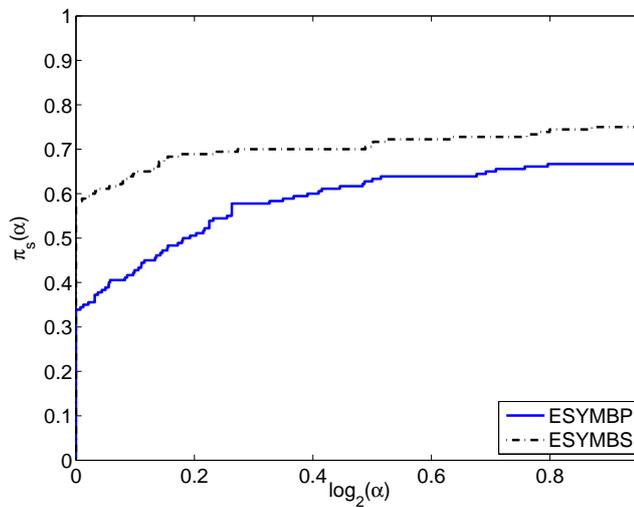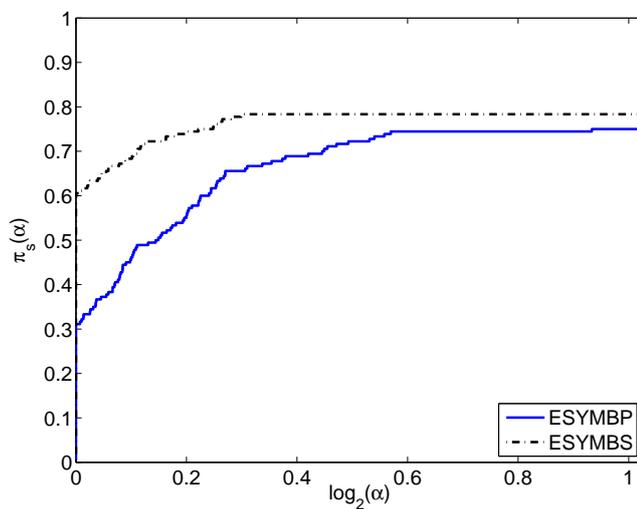

Fig. 4.1: Performance Profile of mean(#F) (RHOEND$= 10^{-2}$)

4.4 A Remark on Extended Symmetric Broyden Update

It is interesting to ask whether the extended symmetric Broyden update can bring improvements to the original version in derivative-free optimization. Powell [41] studies this problem, and it turns out that the extended version rarely reduced the number of function evaluations. We have to point out that it is also the case when σ is selected with our method. However, some numerical experiments of Powell [41] suggested that the extended version might help to improve the convergence rate. It would be desirable to theoretically investigate the local convergence properties of the updates. We think it is still too early to draw any conclusion on the extended symmetric Broyden update, unless we could find strong theoretical evidence.
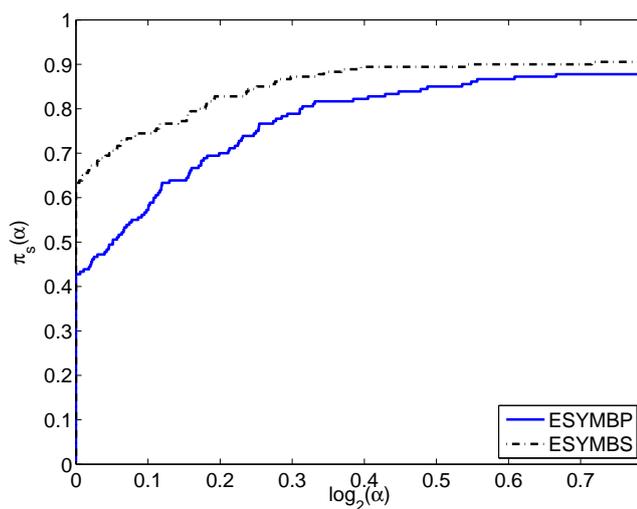
Fig. 4.2: Performance Profile of mean($\#F$) ($\texttt{RHOEND}= 10^{-4}$)

Fig. 4.3: Performance Profile of mean($\#F$) ($\texttt{RHOEND}= 10^{-6}$)

5 Conclusions

Least-norm interpolation has wide applications in derivative-free optimization. Motivated by its objective function, we have studied the H^1 Sobolev seminorm of quadratic functions. We gives an explicit formula for the H^1 seminorm of quadratic functions over balls. It turns out that least-norm interpolation essentially seeks a quadratic interpolant with minimal H^1 seminorm over a ball. Moreover, we find that the parameters x_0 and σ in the objective function determines the center and radius of the ball, respectively. These insights further our understanding of least-norm interpolation.
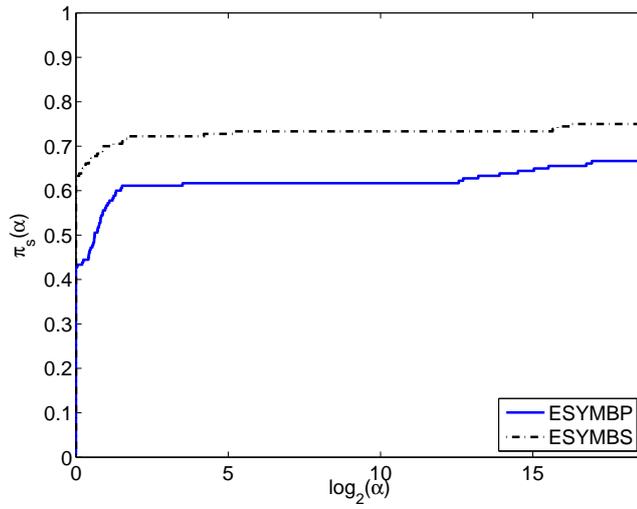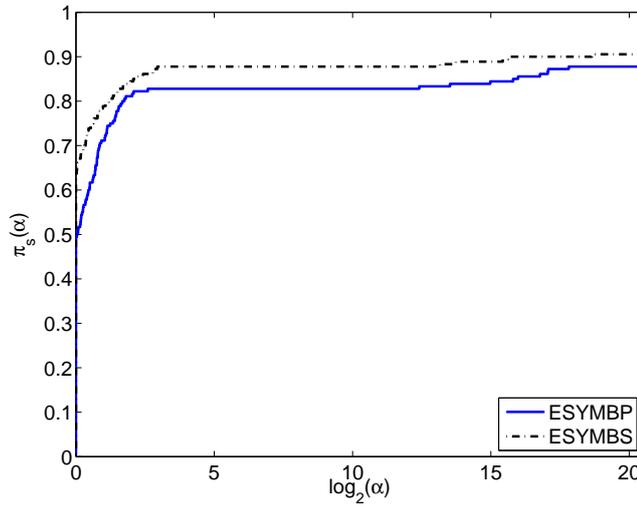
Fig. 4.4: Performance Profile of rstd($\#F$) (RHOEND$= 10^{-2}$)

Fig. 4.5: Performance Profile of rstd($\#F$) (RHOEND$= 10^{-4}$)

With the help of the new observation, we have studied the extended symmetric Broyden update (4.1) proposed by Powell [41]. Since the update calculates the change to the old model by a least-norm interpolation, we can interpret it with our new theory. We have discussed how to choose x_0 and σ for the update according to their geometrical meaning. This discussion leads to the same x_0 used by Powell [41], and a very easy way to choose σ. According to our numerical results, the new σ works better than the one proposed by Powell [41].
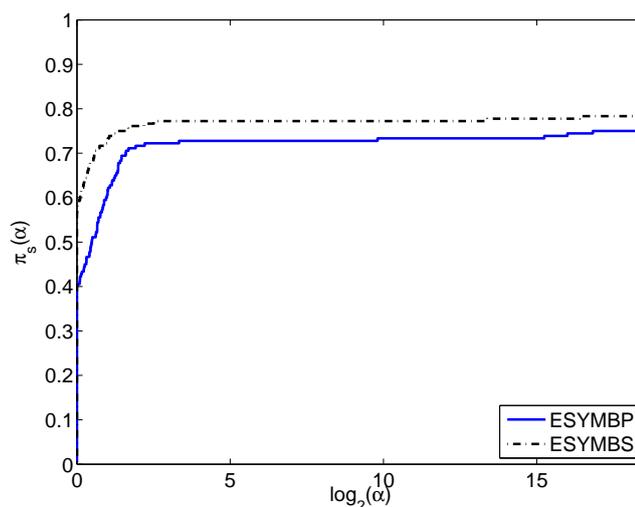
Fig. 4.6: Performance Profile of $\mathrm{rstd}(\#F)$ ($\texttt{RHOEND}= 10^{-6}$)

Our method of designing numerical tests seems reasonable. It provides a more reliable measurement of computation cost than simply counting the numbers of function evaluations, and meanwhile reflects the stability of solvers with respect to computer rounding errors. We hope this method can lead to better benchmarking of derivative-free solvers.

References

1. Adams, R.A.: Sobolev spaces. Pure and applied mathematics. Academic Press (1975)
2. Bandeira, A.S., Scheinberg, K., Vicente, L.N.: Computation of sparse low degree interpolating polynomials and their application to derivative-free optimization. Mathematical Programming **134**(1), 223–257 (2012)
3. Booker, A.J., Dennis, J.E., Frank, P.D., Serafini, D.B., Torczon, V., Trosset, M.W.: A rigorous framework for optimization of expensive functions by surrogates. Structural and Multidisciplinary Optimization **17**(1), 1–13 (1999)
4. Booker, A.J., Dennis Jr, J.E., Frank, P., Serafini, D.B., Torczon, V., Trosset, M.W.: Optimization using surrogate objectives on a helicopter test example. PROGRESS IN SYSTEMS AND CONTROL THEORY **24**, 49–58 (1998)
5. Booker, A.J., Frank, P., Dennis Jr, J.E., Moore, D.W., Serafini, D.B.: Managing surrogate objectives to optimize helicopter rotor design–further experiments. In: IN PROCEEDINGS OF THE SEVENTH AIAA/USAF/NASA/ISSMO SYMPOSIUM ON MULTIDISCIPLINARY ANALYSIS AND OPTIMIZATION (1998)
6. Brent, R.P.: Algorithms for minimization without derivatives. Prentice-Hall, Englewood Cliffs, New Jersey (1973)

7. Choi, T.D., Kelley, C.T.: Superlinear convergence and implicit filtering. SIAM Journal on Optimization **10**(4), 1149–1162 (2000)
8. Conn, A.R., Gould, N.I.M., Toint, Ph.L.: Trust-region methods, *MPS-SIAM series on optimization*, vol. 1. Society for Industrial Mathematics (2000)
9. Conn, A.R., Scheinberg, K., Toint, Ph.L.: On the Convergence of Derivative-Free Methods for Unconstrained Optimization. In: M.D. Buhmann, A. Iserles (eds.) Approximation Theory and Optimization: Tributes to M. J. D. Powell, pp. 83–108. Cambridge University Press, Cambridge, United Kingdom (1997)
10. Conn, A.R., Scheinberg, K., Toint, Ph.L.: Recent progress in unconstrained nonlinear optimization without derivatives. Mathematical Programming **79**, 397–414 (1997)
11. Conn, A.R., Scheinberg, K., Toint, Ph.L.: A derivative free optimization algorithm in practice. In: Proceedings of 7th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization. St. Louis, MO (1998)
12. Conn, A.R., Scheinberg, K., Vicente, L.N.: Introduction to Derivative-Free Optimization. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA (2009)
13. Conn, A.R., Toint, Ph.L.: An Algorithm Using Quadratic Interpolation for Unconstrained Derivative Free Optimization. In: G. Di Pillo, F. Giannessi (eds.) Nonlinear Optimization and Applications, pp. 27–47. Kluwer Academic/Plenum Publishers, New York (1996)
14. Custódio, A.L., Rocha, H., Vicente, L.N.: Incorporating minimum frobenius norm models in direct search. Computational Optimization and Applications **46**(2), 265–278 (2010)
15. Dennis Jr, J.E., Schnabel, R.B.: Least change secant updates for quasi-newton methods. Siam Review pp. 443–459 (1979)
16. Diniz-Ehrhardt, M.A., Martínez, J.M., Raydán, M.: A derivative-free nonmonotone line-search technique for unconstrained optimization. Journal of Computational and Applied Mathematics **219**(2), 383–397 (2008)
17. Dolan, E.D., Moré, J.J.: Benchmarking optimization software with performance profiles. Mathematical Programming **91**(2), 201–213 (2002)
18. Evans, L.C.: Partial differential equations. Graduate studies in mathematics. American Mathematical Society (1998)
19. Fletcher, R.: Practical Methods of Optimization, second edn. John Wiley & Sons, New York (1987)
20. Fowler, K.R., Reese, J.P., Kees, C.E., Dennis Jr, J.E., Kelley, C.T., Miller, C.T., Audet, C., Booker, A.J., Couture, G., Darwin, R.W.: Comparison of derivative-free optimization methods for groundwater supply and hydraulic capture community problems. Advances in Water Resources **31**(5), 743–757 (2008)
21. Gilmore, P., Kelley, C.T.: An implicit filtering algorithm for optimization of functions with many local minima. SIAM Journal on Optimization **5**, 269 (1995)
22. Gould, N.I.M., Orban, D., Toint, Ph.L.: CUTEr and SifDec: A constrained and unconstrained testing environment, revisited. ACM Transactions on Mathematical Software (TOMS) **29**(4), 373–394 (2003)
23. Hemker, T., Fowler, K., Von Stryk, O.: Derivative-free optimization methods for handling fixed costs in optimal groundwater remediation design. In: Proc. of the CMWR XVI-Computational Methods in Water Resources, pp. 19–22 (2006)
24. Kelley, C.: Iterative methods for optimization, *Frontiers in applied mathematics*, vol. 18. Society for Industrial Mathematics (1999)
25. Kelley, C.T.: A brief introduction to implicit filtering. North Carolina State Univ., Raleigh, NC, CRSC Tech. Rep. CRSC-TR02-28 (2002)
26. Kelley, C.T.: Implicit filtering. Society for Industrial and Applied Mathematics (2011)
27. Kolda, T.G., Lewis, R.M., Torczon, V.: Optimization by direct search: New perspectives on some classical and modern methods. Siam Review pp. 385–482 (2003)
28. Lewis, R.M., Torczon, V., Trosset, M.W.: Direct search methods: then and now. Journal of Computational and Applied Mathematics **124**(1), 191–207 (2000)
29. Marazzi, M., Nocedal, J.: Wedge trust region methods for derivative free optimization. Mathematical programming **91**(2), 289–305 (2002)
30. Oeuvray, R.: Trust-region methods based on radial basis functions with application to biomedical imaging. Ph.D. thesis, ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE (2005)
31. Oeuvray, R., Bierlaire, M.: A new derivative-free algorithm for the medical image registration problem. International Journal of Modelling and Simulation **27**(2), 115–124 (2007)
32. Oeuvray, R., Bierlaire, M.: a derivative-free algorithm based on radial basis functions. In: International Journal of Modelling and Simulation (2008)
33. Powell, M.J.D.: A direct search optimization method that models the objective and constraint functions by linear interpolation. Advances in Optimization and Numerical Analysis pp. 51–67 (1994)
34. Powell, M.J.D.: Direct search algorithms for optimization calculations. Acta Numerica **7**(1), 287–336 (1998)

35. Powell, M.J.D.: UOBYQA: unconstrained optimization by quadratic approximation. Tech. Rep. DAMTP NA2000/14, CMS, University of Cambridge (2000)
36. Powell, M.J.D.: On trust region methods for unconstrained minimization without derivatives. Mathematical programming **97**(3), 605–623 (2003)
37. Powell, M.J.D.: Least frobenius norm updating of quadratic models that satisfy interpolation conditions. Mathematical Programming **100**, 183–215 (2004)
38. Powell, M.J.D.: The NEWUOA software for unconstrained optimization without derivatives. Tech. Rep. DAMTP NA2004/08, CMS, University of Cambridge (2004)
39. Powell, M.J.D.: Developments of NEWUOA for minimization without derivatives. IMA Journal of Numerical Analysis pp. 649–664 (2008). DOI 10.1093/imanum/drm047
40. Powell, M.J.D.: The BOBYQA algorithm for bound constrained optimization without derivatives. Tech. Rep. DAMTP 2009/NA06, CMS, University of Cambridge (2009)
41. Powell, M.J.D.: Beyond symmetric Broyden for updating quadratic models in minimization without derivatives. Mathematical Programming pp. 1–26 (2012). DOI 10.1007/s10107-011-0510-y. URL http://dx.doi.org/10.1007/s10107-011-0510-y
42. Stewart III, G.: A modification of davidon's minimization method to accept difference approximations of derivatives. Journal of the ACM (JACM) **14**(1), 72–83 (1967)
43. Vanden Berghen, F., Bersini, H.: CONDOR, a new parallel, constrained extension of powell's UOBYQA algorithm: Experimental results and comparison with the DFO algorithm. Journal of Computational and Applied Mathematics **181**(1), 157–175 (2005)
44. Wild, S.M.: MNH: a derivative-free optimization algorithm using minimal norm Hessians. In: Tenth Copper Mountain Conference on Iterative Methods (2008)
45. Wild, S.M., Regis, R.G., Shoemaker, C.A.: ORBIT: Optimization by radial basis function interpolation in trust-regions. SIAM Journal on Scientific Computing **30**(6), 3197–3219 (2008)
46. Winfield, D.H.: Function minimization by interpolation in a data table. IMA Journal of Applied Mathematics **12**(3), 339 (1973)
47. Wright, M.H.: Direct search methods: Once scorned, now respectable. PITMAN RESEARCH NOTES IN MATHEMATICS SERIES pp. 191–208 (1996)