Globally convergent algorithms for finding zeros of duplomonotone mappings

Francisco J. Aragón Artacho[*]   Ronan M.T. Fleming[†]

September 1, 2014

Systems Biochemistry Group, Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Campus Belval, 4362 Esch-sur-Alzette, Luxembourg.

Abstract

We introduce a new class of mappings, called duplomonotone, which is strictly broader than the class of monotone mappings. We study some of the main properties of duplomonotone functions and provide various examples, including nonlinear duplomonotone functions arising from the study of systems of biochemical reactions. Finally, we present three variations of a derivative-free line search algorithm for finding zeros of systems of duplomonotone equations, and we prove their linear convergence to a zero of the function.

Keywords: Generalized monotonicity, duplomonotone mapping, monotone mapping, global convergence, line search method, derivative-free algorithm, biochemical reactions.

1   Introduction

Monotone mappings have been extensively studied in the literature, see for instance [6, Chapter 12] or the recent monograph [1]. In many practical problems, though, the monotonicity assumption turns out to be too strong. Consequently, several generalized notions of monotonocity have been introduced and thoroughly studied by various authors in order to relax it while keeping some of the useful properties of monotone mappings, see [2, 4] and the references therein.

In mathematical models of biochemical reaction networks [3], a problem arises of finding a zero of functions that are typically not monotone (see Example 2.14). These functions seem to have a *generalized* monotonicity property that has not yet appeared in the literature but can be exploited to find a zero of such functions. In this paper we introduce this new class of generalized monotone mappings, which we call *duplomonotone*, and present a rather simple derivative-free line search algorithm that can be used to find a zero of a duplomonotone function.

The paper is organized as follows: in Section 2 we introduce duplomonotone mappings, analyze their basic properties and provide various illustrative examples; in Section 3 we present three variations of a derivative-free line search algorithm for finding a zero of a duplomonotone function, and we prove their linear convergence under strong duplomonotonicity plus some Lipschitz-type assumption on the points of the lower level set defined by the initial point.

Throughout, $\|\cdot\|$ denotes the Euclidean norm, while the usual inner product is denoted by $\langle\cdot,\cdot\rangle$. We say that F is a *set-valued mapping* from $\mathbb{R}^m$ to $\mathbb{R}^n$, denoted by $F:\mathbb{R}^m \rightrightarrows \mathbb{R}^n$, if for every $x \in \mathbb{R}^m$, $F(x)$ is a subset of $\mathbb{R}^n$. The gradient of a differentiable function $f:\mathbb{R}^m \to \mathbb{R}^n$ at some point $x \in \mathbb{R}^m$ is denoted by $\nabla f(x) \in \mathbb{R}^{m \times n}$.

---

[*]E-mail: francisco.aragon@ua.es

[†]E-mail: ronan.mt.fleming@gmail.com

2 Duplomonotonicity

Recall that a function $f : \mathbb{R}^m \to \mathbb{R}^m$ is said to be *monotone* when

$$\langle f(x) - f(y), x - y \rangle \geq 0 \quad \text{for all } x, y \in \mathbb{R}^m,$$

and *strictly monotone* if this inequality is strict whenever $x \neq y$. Further, f is called *strongly monotone* for some $\sigma > 0$ when

$$\langle f(x) - f(y), x - y \rangle \geq \sigma \|x - y\|^2 \quad \text{for all } x, y \in \mathbb{R}^m.$$

We introduce next a new property that is implied by monotonicity.

Definition 2.1. A function $f : \mathbb{R}^m \to \mathbb{R}^m$ is called *duplomonotone* with constant $\bar{\tau} > 0$ if

$$\langle f(x) - f(x - \tau f(x)), f(x) \rangle \geq 0 \quad \text{whenever } x \in \mathbb{R}^m, 0 \leq \tau \leq \bar{\tau}, \tag{1}$$

and *strictly duplomonotone* if this inequality is strict whenever $f(x) \neq 0$. The function f is said to be *strongly duplomonotone* for some $\sigma > 0$ with constant $\bar{\tau} > 0$ if

$$\langle f(x) - f(x - \tau f(x)), f(x) \rangle \geq \sigma\tau \|f(x)\|^2 \quad \text{whenever } x \in \mathbb{R}^m, 0 \leq \tau \leq \bar{\tau}. \tag{2}$$

The *modulus of strong duplomonotonicity* is the supremum of the constants σ for which (2) holds.

Remark 2.2. Letting σ be zero in (2) will allow us to handle both duplomonotonicity and strong duplomonotonicity at the same time. Hence, we refer to this as f being strongly duplomonotone with $\sigma \geq 0$.

Obviously, every (strongly) monotone function is (strongly) duplomonotone. In the next simple example we show that the converse is not true in general: the class of duplomonotone functions is strictly broader than the class of monotone functions. Thus, we have:

$$\boxed{\text{monotonicity} \quad \genfrac{}{}{0pt}{}{\Rightarrow}{\nLeftarrow} \quad \text{duplomonotonicity}}$$

Example 2.3. Given a matrix $A \in \mathbb{R}^{m \times m}$, consider the linear function $f(x) := Ax$. Recall that the *symmetric part* of A is the matrix $A_s := \frac{1}{2}(A + A^T)$. The mapping f is monotone if and only if A_s is positive semidefinite (see e.g. [6, Example 12.2]). On the other hand, f is duplomonotone if and only if there is some $\bar{\tau} > 0$ such that, for any $x \in \mathbb{R}^m$, one has

$$0 \leq \langle f(x) - f(x - \tau f(x)), f(x) \rangle = \tau x^T A^T A^2 x, \quad \text{whenever } 0 \leq \tau \leq \bar{\tau};$$

that is, f is duplomonotone if and only if $(A^T A^2)_s$ is positive semidefinite. Furthermore, f is strongly duplomonotone for $\sigma > 0$ if and only if for any $x \in \mathbb{R}^m$ and any positive τ, one has

$$
\begin{aligned}
0 \leq \langle f(x) - f(x - \tau f(x)), f(x) \rangle - \sigma\tau \|f(x)\|^2 &= \tau x^T A^T A^2 x - \sigma\tau x^T A^T A x \\
&= \tau x^T A^T (A - \sigma I) A x,
\end{aligned}
$$

where I denotes the identity mapping. Therefore, f is strongly duplomonotone for $\sigma > 0$ if and only if $(A^T(A - \sigma I)A)_s = A^T(A_s - \sigma I)A$ is positive semidefinite.

If A is symmetric, then $(A^T A^2)_s = A^3$, whose eigenvalues have the same sign as those of A. Thus, for A symmetric, the function f is duplomonotone if and only if f is monotone. However, this may not be the case if A is asymmetric. As a simple example, if we take

$$A := \begin{bmatrix} 2 & 0 \\ 2 & 0 \end{bmatrix}, \tag{3}$$

then,

$$A_s = \begin{bmatrix} 2 & 1 \\ 1 & 0 \end{bmatrix},$$

which is not positive semidefinite, while

$$\left(A^T A^2\right)_s = \begin{bmatrix} 16 & 0 \\ 0 & 0 \end{bmatrix}$$

is positive semidefinite. Thus, $f((x,y)^T) = (2x, 2x)^T$ is duplomonotone but is not monotone. Moreover, it is not difficult to check that f is not even *quasimonotone*[1]. In fact, f is strongly duplomonotone with modulus $\sigma = 2$. Indeed,

$$A^T(A - \sigma I)A = \begin{bmatrix} 8(2 - \sigma) & 0 \\ 0 & 0 \end{bmatrix},$$

which is positive semidefinite if and only if $\sigma \leq 2$.

A strictly monotone function has at most one zero. This is not the case for duplomonotone functions: even under strong duplomonotonicity we can see that the function $f(x) = Ax$ with A given by (3) has a zero at $(0, y)^T$ for every $y \in \mathbb{R}$. In fact, the zero function is strongly duplomonotone for any $\sigma > 0$.

We have shown a function in Example 2.3 that is duplomonotone but not quasimonotone. It is interesting to note that there are also functions that are quasimonotone but not duplomonotone, e.g. $f(x) = -|x|$ for $x \in \mathbb{R}$.

Example 2.4. Given a matrix $A \in \mathbb{R}^{m \times m}$ and a vector $b \in \mathbb{R}^m$, consider the affine function $f(x) := Ax + b$. By [6, Example 12.2], f is monotone if and only if A_s is positive semidefinite. On the other hand, f is duplomonotone if and only if

$$(Ax + b)^T A(Ax + b) \geq 0 \quad \text{for all } x \in \mathbb{R}^m;$$

that is, f is duplomonotone if and only if A_s is positive semidefinite on the range of f. For example, one can check that for A given in (3) and any $b = (b_1, b_2)^T \in \mathbb{R}^2$, the function f is duplomonotone if and only if $b_1 = b_2$.

Next we present a direct characterization of duplomonotonicity in terms of the Euclidean norm.

Proposition 2.5. *A function $f : \mathbb{R}^m \to \mathbb{R}^m$ is strongly duplomonotone for $\sigma \geq 0$ if and only if there is some constant $\bar{\tau} > 0$ such that for all $x \in \mathbb{R}^m$ and all $0 \leq \tau \leq \bar{\tau}$ one has*

$$\|f(x - \tau f(x))\|^2 \leq (1 - 2\sigma\tau)\|f(x)\|^2 + \|f(x - \tau f(x)) - f(x)\|^2. \tag{4}$$

Proof. For any $x \in \mathbb{R}^m$ and any $\tau > 0$, we have.

$$\begin{aligned} \|f(x - \tau f(x))\|^2 &= \|(f(x - \tau f(x)) - f(x)) + f(x)\|^2 \\ &= \|f(x - \tau f(x)) - f(x)\|^2 + \|f(x)\|^2 + 2\langle f(x - \tau f(x)) - f(x), f(x)\rangle. \end{aligned}$$

The stated equivalence follows then from the definition of strong duplomonotonicity of f. $\qquad\square$

The following example shows the importance of considering the constant $\bar{\tau}$ in the definition of duplomonotonicity: there are functions for which (1) does not hold for all $\tau > 0$. One could also define a weaker notion of duplomonotonicity where the constant $\bar{\tau}$ in (1) depends on each point x. Nevertheless, this property might be too weak to guarantee the convergence of the line search algorithms in Section 3, as we need to ensure that the step size is bounded away from zero.

---

[1] A function $f : \mathbb{R}^m \to \mathbb{R}^m$ is *quasimonotone* if the following implication holds:

$$\langle f(x), y - x\rangle > 0 \Rightarrow \langle f(y), y - x\rangle \geq 0,$$

for every $x, y \in \mathbb{R}^m$. Monotonicity implies quasimonotonicity.

Example 2.6. Let $f : \mathbb{R}^2 \to \mathbb{R}^2$ be given by $f(x) := (x_1 x_2^2, x_2)^T$ for $x := (x_1, x_2)^T \in \mathbb{R}^2$. It is easy to check that f is not monotone: if we take $x := (-3, 0)^T$ and $y := (-1, 1)^T$, we have

$$\langle f(x) - f(y), x - y \rangle = -1.$$

On the other hand, after some algebraic manipulation, one can show that for all $x := (x_1, x_2)^T \in \mathbb{R}^2$, one has

$$\langle f(x) - f(x - \tau f(x)), f(x) \rangle = \left((\tau - 1)^2 x_1^2 x_2^4 + (2 - \tau) x_1^2 x_2^2 + 1 \right) \tau x_2^2,$$

which is nonnegative for all $\tau \in [0, 2]$. Thus, f is duplomonotone with constant $\bar{\tau} = 2$. If $\tau > 2$, the expression above can be negative for some $x \in \mathbb{R}^2$. Indeed, choose any $\varepsilon > 0$ and let $z := (z_1, \sqrt{\varepsilon/2}/(\varepsilon + 1))$ for some $z_1 \in \mathbb{R}$. Then,

$$\langle f(z) - f(z - (2 + \varepsilon)f(z)), f(z) \rangle = \frac{-\left(\varepsilon^4 + 2\varepsilon^3 \right) z_1^2 + 4\varepsilon^4 + 16\varepsilon^3 + 20\varepsilon^2 + 8\varepsilon}{8(\varepsilon + 1)^4},$$

which is negative for z_1^2 sufficiently big.

The next result shows that if a function is both Lipschitz continuous and strongly duplomonotone for $\sigma > 0$, then σ is bounded above by the Lipschitz constant.

Proposition 2.7. *If a function $f : \mathbb{R}^m \to \mathbb{R}^m$ is Lipschitz continuous with constant $\ell > 0$ and strongly duplomonotone for $\sigma > 0$, with $f \not\equiv 0$, then $\sigma \leq \ell$.*

Proof. Because of the Lipschitz continuity, we have

$$\| f(x - \tau f(x)) - f(x) \| \leq \ell \tau \| f(x) \| \quad \text{for all } x \in \mathbb{R}^m, \tau > 0.$$

Let $\bar{\tau} > 0$ be the strong duplomonotonicity constant in (2), and pick any $z \in \mathbb{R}^m$ such that $f(z) \neq 0$. Then

$$\sigma \bar{\tau} \| f(z) \|^2 \leq \langle f(z) - f(z - \bar{\tau} f(z)), f(z) \rangle \leq \| f(z - \bar{\tau} f(z)) - f(z) \| \| f(z) \| \leq \ell \bar{\tau} \| f(z) \|^2,$$

whence $\sigma \leq \ell$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

In the following result we show a direct consequence of duplomonotonicity for differentiable functions.

Proposition 2.8. *Let $f : \mathbb{R}^m \to \mathbb{R}^m$ be differentiable. The following assertions hold.*

(i) *If f is duplomonotone, then*

$$f(x)^T \nabla f(x) f(x) \geq 0 \quad \text{for all } x \in \mathbb{R}^m. \tag{5}$$

(ii) *If f is strongly duplomonotone for $\sigma > 0$, then*

$$f(x)^T \nabla f(x) f(x) \geq \sigma \| f(x) \|^2 \quad \text{for all } x \in \mathbb{R}^m. \tag{6}$$

Proof. Assume that f satisfies (2) with $\sigma \geq 0$ and $\bar{\tau} > 0$. Fix $x \in \mathbb{R}^m$ and choose an arbitrary $\tau \in (0, \bar{\tau}]$. Dividing (2) by τ we get

$$-\left\langle \frac{f(x - \tau f(x)) - f(x)}{\tau}, f(x) \right\rangle \geq \sigma \| f(x) \|^2,$$

and taking the limit as $\tau \searrow 0$, we obtain $f(x)^T \nabla f(x) f(x) \geq \sigma \| f(x) \|^2$. $\qquad\qquad\qquad\square$

Remark 2.9. (i) In general, strict duplomonotonicity does not imply that equality in (5) is only achieved when $f(x) = 0$, in the same way that strict monotonicity does not imply positive definiteness of $\nabla f(x)$.
(ii) Observe that both assertions also hold under the weaker notion of duplomonotonicity where the constant $\bar{\tau}$ depends on each $x \in \mathbb{R}^m$.

For differentiable functions in one dimension, the notions of (strong) duplomonotonicity and (strong) monotonicity agree. In fact, we will prove in Proposition 2.11 that the concepts of monotonicity and duplomonotonicity coincide for continuous functions in one dimension[2]. This is not the case if the function is not continuous, as we show in Example 2.12.

Corollary 2.10. *Let $f : \mathbb{R} \to \mathbb{R}$ be differentiable. Then f is (strongly) monotone if and only if f is (strongly) duplomonotone.*

Proof. This is just a consequence of Proposition 2.8 and the fact that f is (strongly) monotone with constant $\sigma \geq 0$ if and only if $f'(x) \geq \sigma$. $\qquad\square$

Proposition 2.11. *Let $f : \mathbb{R} \to \mathbb{R}$ be continuous. Then f is monotone if and only if f is duplomonotone.*

Proof. Suppose that f is duplomonotone with constant $\bar{\tau} > 0$. If there is some $z \in \mathbb{R}$ such that $f(z) > 0$, then we claim that there is an open interval containing z such that $f(z)$ is both nondecreasing and positive on it. Indeed, by continuity of f, there is some $\delta_0 > 0$ such that $f(x) > f(z)/2 > 0$ for all $x \in (z - \delta_0, z + \delta_0)$. Set $\delta := \min\{\delta_0, \bar{\tau}f(z)/4\}$. Choose any $x, y \in (z - \delta, z + \delta)$ with $x > y$, and set $\tau := \frac{x-y}{f(x)} \in (0, \bar{\tau})$. Then, $x - \tau f(x) = y$. From the duplomonotonicity of f, we deduce
$$0 \leq f(x) - f(x - \tau f(x)) = f(x) - f(y).$$
Hence, f is nondecreasing and positive on $(z - \delta, z + \delta)$, as claimed.

Observe now that f has to be positive and nondecreasing on $(z - \delta, +\infty)$, again by continuity of f. Therefore, if we set $a := \inf\{x \in \mathbb{R} \mid f(x) > 0\} \in \mathbb{R} \cup \{-\infty, +\infty\}$, it follows that $\{x \in \mathbb{R} \mid f(x) > 0\} = (a, +\infty)$ and f is nondecreasing on $(a, +\infty)$. Using the same argument, we deduce that $\{x \in \mathbb{R} \mid f(x) < 0\} = (-\infty, b)$ with $b \in \mathbb{R} \cup \{-\infty, +\infty\}$ and f is nondecreasing on $(-\infty, b)$. Thus, f is monotone. $\qquad\square$

Example 2.12. Consider the function $f : \mathbb{R} \to \mathbb{R}$ defined for $x \in \mathbb{R}$ by
$$f(x) := \begin{cases} 0, & \text{if } x \in \mathbb{Q}; \\ 1, & \text{if } x \notin \mathbb{Q}. \end{cases}$$

The function f is not monotone (not even locally):
$$(f(\pi) - f(4))(\pi - 4) = \pi - 4 < 0.$$

On the other hand, f is duplomonotone: for any $x \in \mathbb{Q}$ the duplomonotonicity condition (1) trivially holds since $f(x) = 0$, while for any $x \notin \mathbb{Q}$ and any $\tau > 0$ we have
$$(f(x) - f(x - \tau f(x)))\, f(x) = 1 - f(x - \tau) \geq 0.$$

Furthermore, one can easily check that this function is not strongly duplomonotone. A slight modification of this example yields a function that is strongly duplomonotone, but still not monotone: let $g : \mathbb{R} \to \mathbb{R}$ be defined for $x \in \mathbb{R}$ by
$$g(x) := \begin{cases} 0, & \text{if } x \in \mathbb{Q}; \\ x, & \text{if } x \notin \mathbb{Q}. \end{cases}$$

Again, the function g is not monotone (not even locally), since
$$(g(\pi) - g(4))(\pi - 4) = \pi(\pi - 4) < 0.$$

---

[2] This result and the proof included here is due to the referee of this paper, who noticed that the Dirichlet function in Example 2.12 is not monotone because it is not continuous.

In this case, g is strongly duplomonotone for $\sigma = 1$ with constant $\bar{\tau} = 1$: for any $x \notin \mathbb{Q}$ and any $\tau \in [0,1]$ we have

$$
\begin{aligned}
(g(x) - g(x - \tau g(x)))\, g(x) - \tau g(x)^2 &= (x - g((1-\tau)x))x - \tau x^2 \\
&= \begin{cases} (1-\tau)x^2, & \text{if } (1-\tau)x \in \mathbb{Q} \\ 0, & \text{if } (1-\tau)x \notin \mathbb{Q} \end{cases} \\
&\geq 0.
\end{aligned}
$$

Therefore, without differentiability, the concepts of monotonicity and duplomonotonicity may be quite different, even in one dimension.

In the next proposition we introduce a property that implies duplomonotonicity, but is still weaker than monotonicity (see Example 2.14). This property has a characterization for differentiable functions analogous to the positive-semidefiniteness of the Jacobian for monotone functions, see e.g. [6, Proposition 12.3].

Proposition 2.13. *Let $f : \mathbb{R}^m \to \mathbb{R}^m$ be differentiable. Then, for any $\sigma \geq 0$, the following two properties are equivalent:*

(i) $\langle f(x - \tau_1 f(x)) - f(x - \tau_2 f(x)), f(x) \rangle \geq \sigma(\tau_2 - \tau_1)\|f(x)\|^2$ *for all $x \in \mathbb{R}^m$, $0 \leq \tau_1 \leq \tau_2 \leq \bar{\tau}$;*

(ii) $f(x)^T \nabla f(x - \tau f(x)) f(x) \geq \sigma \|f(x)\|^2$ *for all $x \in \mathbb{R}^m$, $\tau \in [0, \bar{\tau}]$.*

Proof. Assume that (i) holds. Choose any $x \in \mathbb{R}^m$ and any $\tau \in [0, \bar{\tau})$. For all $t \in (0, \bar{\tau} - \tau]$ one has

$$
-\langle f(x - (t + \tau)f(x)) - f(x - \tau f(x)), f(x) \rangle \geq \sigma t \|f(x)\|^2.
$$

Thus, dividing by t and taking the limit as $t \searrow 0$,

$$
\sigma \|f(x)\|^2 \leq -\left\langle \lim_{t \searrow 0} \frac{f(x - (t+\tau)f(x)) - f(x - \tau f(x))}{t}, f(x) \right\rangle = \langle \nabla f(x - \tau f(x))^T f(x), f(x) \rangle,
$$

so (ii) follows.

Conversely, assume that (ii) holds. Pick any $x \in \mathbb{R}^m$ and any $0 \leq \tau_1 \leq \tau_2 \leq \bar{\tau}$. Consider the function

$$
h(\lambda) := \langle f\left(x - (\lambda \tau_1 + (1-\lambda)\tau_2)f(x)\right) - f(x - \tau_2 f(x)) - \sigma \lambda(\tau_2 - \tau_1)f(x), f(x) \rangle
$$

for $\lambda \in \mathbb{R}$. Then, by (ii),

$$
h'(\lambda) = \langle \nabla f\left(x - (\lambda\tau_1 + (1-\lambda)\tau_2)f(x)\right)^T (\tau_2 - \tau_1)f(x) - \sigma(\tau_2 - \tau_1)f(x), f(x) \rangle \geq 0,
$$

for all $\lambda \in [0,1]$, whence,

$$
0 = h(0) \leq h(1) = \langle f(x - \tau_1 f(x)) - f(x - \tau_2 f(x)) - \sigma(\tau_2 - \tau_1)f(x), f(x) \rangle,
$$

which implies (i). $\qquad\square$

Our motivation to characterize duplomonotone mappings arose from mathematical modeling of networks of (bio)chemical reactions, an increasingly prominent application of mathematical and numerical optimization. The next example introduces a very simple (bio)chemical reaction network, involving three molecules and three reactions, where each row of x corresponds to the logarithmic abundance of some molecule and each row of $-f(x)$ corresponds to the rate of change of abundance per unit time.

Example 2.14. Consider the function $f : \mathbb{R}^3 \to \mathbb{R}^3$ defined for $x \in \mathbb{R}^3$ by $f(x) := ([F, R] - [R, F]) \exp([F, R]^T x)$, where $\exp(\cdot)$ denotes the component-wise exponential,

$$
F := \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, R := \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},
$$

and $[\,\cdot\,,\cdot\,]$ is the horizontal concatenation operator. That is, for $x := (x_1, x_2, x_3)^T \in \mathbb{R}^3$ we have

$$f(x) = \begin{bmatrix} 2e^{x_1} - e^{x_2} - e^{x_3} \\ -e^{x_1} + 2e^{x_2} - e^{x_3} \\ -e^{x_1} - e^{x_2} + 2e^{x_3} \end{bmatrix}.$$

The function f is not monotone because $\nabla f(x)$ is not positive semidefinite for all $x \in \mathbb{R}^3$. For instance, if $z := (0, 0, \log(2))^T$ and $w := (3, 3, 2)^T$, we have

$$w^T \nabla f(z) w = -2.$$

Nevertheless, the function f is duplomonotone because, in fact, it satisfies Proposition 2.13(ii) with $\sigma = 0$. Indeed, if we define

$$\varphi(x, \tau) := \langle f(x) - f(x - \tau f(x)), f(x) \rangle,$$

we have

$$\frac{\partial \varphi}{\partial \tau}(x, \tau) = \langle \nabla f(x - \tau f(x))^T f(x), f(x) \rangle. \tag{7}$$

After some algebraic manipulation, we obtain

$$\begin{aligned}
\varphi(x, \tau) &= 3e^{x_1 + \tau(-2e^{x_1} + e^{x_2} + e^{x_3})}(-2e^{x_1} + e^{x_2} + e^{x_3}) \\
&\quad + 3e^{x_2 + \tau(e^{x_1} - 2e^{x_2} + e^{x_3})}(e^{x_1} - 2e^{x_2} + e^{x_3}) \\
&\quad + 3e^{x_3 + \tau(e^{x_1} + e^{x_2} - 2e^{x_3})}(e^{x_1} + e^{x_2} - 2e^{x_3}) \\
&\quad + (-2e^{x_1} + e^{x_2} + e^{x_3})^2 + (e^{x_1} - 2e^{x_2} + e^{x_3})^2 \\
&\quad + (e^{x_1} + e^{x_2} - 2e^{x_3})^2.
\end{aligned}$$

Thus,

$$\begin{aligned}
\frac{\partial \varphi}{\partial \tau}(x, \tau) &= 3e^{x_1 + \tau(-2e^{x_1} + e^{x_2} + e^{x_3})}(-2e^{x_1} + e^{x_2} + e^{x_3})^2 \\
&\quad + 3e^{x_2 + \tau(e^{x_1} - 2e^{x_2} + e^{x_3})}(e^{x_1} - 2e^{x_2} + e^{x_3})^2 \\
&\quad + 3e^{x_3 + \tau(e^{x_1} + e^{x_2} - 2e^{x_3})}(e^{x_1} + e^{x_2} - 2e^{x_3})^2 \\
&\geq 0,
\end{aligned}$$

and because of (7), we have that Proposition 2.13(ii) holds for all $\tau > 0$.

Indeed, the function f is strictly duplomonotone because $\frac{\partial \varphi}{\partial \tau}(x, \tau) > 0$ for all $x \notin \Omega$, where

$$\Omega := \left\{ x \in \mathbb{R}^3 \mid f(x) = 0 \right\} = \left\{ (x_1, x_2, x_3)^T \in \mathbb{R}^3 \mid x_1 = x_2 = x_3 \right\}.$$

Hence, $\varphi(x, \tau) > \varphi(x, 0) = 0$ for all $x \notin \Omega$ and all $\tau > 0$; that is, f is strictly duplomonotone.

The sum of two monotone operators is clearly monotone. Further, if a mapping F is monotone, one can easily show that for all $\alpha > 0$ the mapping $F + \alpha I$ is strongly monotone. Do these properties also hold for duplomonotone functions? The answer is negative in general. As we show in the next example, duplomonotonicity can be destroyed by the addition of a monotone linear function of arbitrarily small slope.

Example 2.15. Consider the matrix

$$A := \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

By Example 2.3, the function $f(x) := Ax$ is duplomonotone, since $A^T A^2 = 0_{2 \times 2}$. Choose any $\alpha > 0$ and consider the function $g(x) := Bx$, with $B := A + \alpha I$. Then,

$$\left(B^T B^2 \right)_s = \begin{bmatrix} \alpha^3 & \frac{3}{2}\alpha^2 \\ \frac{3}{2}\alpha^2 & \alpha^3 + 2\alpha \end{bmatrix}.$$

The eigenvalues of $\left(B^T B^2 \right)_s$ are $\alpha^3 + \alpha \pm 1/2\alpha\sqrt{9\alpha^2 + 4}$. If $\alpha \in (0, 1/2)$, we have that $\alpha^3 + \alpha - 1/2\alpha\sqrt{9\alpha^2 + 4} < 0$. Therefore, the function $g = f + \alpha I$ is not duplomonotone for any $\alpha \in (0, 1/2)$.

A direct consequence of Proposition 2.8 is that $-f(x)$ is a descent direction for $\|f(\cdot)\|^2$ at any point $x \in \mathbb{R}^m$ when f is duplomonotone. This property inspires the derivative-free algorithms in Section 3 for finding zeros of the function f.

Corollary 2.16. *Let $f : \mathbb{R}^m \to \mathbb{R}^m$ be differentiable and strongly duplomonotone for $\sigma > 0$. Then, for all $x \in \mathbb{R}^m$, either $f(x) = 0$ or the vector $-f(x)$ provides a descent direction for the merit function $\|f(\cdot)\|^2$ at the point x.*

Proof. Observe that, for any $x \in \mathbb{R}^m$, we have $\nabla \left(\|f(\cdot)\|^2 \right)(x) = 2\nabla f(x)f(x)$. Thus, inequality (6) implies that

$$\langle \nabla \left(\|f(\cdot)\|^2 \right)(x), -f(x) \rangle = -2\langle \nabla f(x)f(x), f(x) \rangle \leq -2\sigma \|f(x)\|^2. \tag{8}$$

The assertion follows. $\qquad \square$

It is straightforward to extend the definition of duplomonotonicity for set-valued mappings.

Definition 2.17. A set-valued mapping $F : \mathbb{R}^m \rightrightarrows \mathbb{R}^m$ is called *duplomonotone* with constant $\bar{\tau} > 0$ if for all $x \in \mathbb{R}^m$ and all $\tau \in [0, \bar{\tau}]$ one has

$$\langle y_0 - y_1, y_0 \rangle \geq 0 \quad \text{whenever } y_0 \in F(x), y_1 \in F(x - \tau y_0).$$

The mapping F is said to be *strongly duplomonotone* for some $\sigma > 0$ with constant $\bar{\tau} > 0$ if for all $x \in \mathbb{R}^m$ and all $\tau \in [0, \bar{\tau}]$ one has

$$\langle y_0 - y_1, y_0 \rangle \geq \sigma\tau \|y_0\|^2 \quad \text{whenever } y_0 \in F(x), y_1 \in F(x - \tau y_0).$$

One can easily extend the characterization of duplomonotonicity given in Proposition 2.5 to set-valued mappings.

Proposition 2.18. *A set-valued mapping $F : \mathbb{R}^m \rightrightarrows \mathbb{R}^m$ is strongly duplomonotone for $\sigma \geq 0$ if and only if there is some $\bar{\tau} > 0$ such that for all $x \in \mathbb{R}^m$ and all $\tau \in [0, \bar{\tau}]$ one has*

$$\|y_1\|^2 \leq (1 - 2\sigma\tau)\|y_0\|^2 + \|y_1 - y_0\|^2 \quad \text{whenever } y_0 \in F(x), y_1 \in F(x - \tau y_0).$$

We will not explore duplomonotone set-valued mappings any further here, as it is beyond the scope of the present paper.

3 Derivative-free algorithms for systems of duplomonotone equations

In this section we consider the problem of finding solutions of systems of nonlinear equations

$$f(x) = 0, \tag{9}$$

where $f : \mathbb{R}^m \to \mathbb{R}^m$ is strongly duplomonotone for $\sigma > 0$. Corollary 2.16 drives us to consider the following derivative-free line search algorithm for finding zeros of f.

Algorithm 1: Backtracking (with $\sigma > 0$ known)

Fix a precision $\varepsilon > 0$. Choose any $x_0 \in \mathbb{R}^m$, $0 < \alpha < 2\sigma$, $0 < \beta < 1$, and set $k := 0$.

while $\|f(x_k)\| > \varepsilon$ **do**

   Compute $\lambda_k := \frac{1}{\alpha}\beta^{p_k}$, where p_k is the smallest positive integer such that

$$\|f(x_k - \lambda_k f(x_k))\|^2 \leq (1 - \alpha\lambda_k)\|f(x_k)\|^2. \tag{10}$$

   $x_{k+1} := x_k - \lambda_k f(x_k)$;
   $k := k + 1$;

end

Observe that, when f is differentiable, the step size acceptance criteria (10) is implied by the usual Armijo rule for the function $\|f(\cdot)\|^2$ and the direction $d_k := -f(x_k)$. Indeed, given some constant $c \in (0,1)$, the Armijo rule for $\|f(\cdot)\|^2$ will search for a step size λ_k satisfying

$$\begin{aligned} \|f(x_k + \lambda_k d_k)\|^2 &\leq & \|f(x_k)\|^2 + 2c\lambda_k d_k^T \nabla f(x_k) f(x_k) \\ &=& \|f(x_k)\|^2 - 2c\lambda_k f(x_k)^T \nabla f(x_k) f(x_k). \end{aligned}$$

Proposition 2.8(ii) gives us

$$\|f(x_k)\|^2 - 2c\lambda_k f(x_k)^T \nabla f(x_k) f(x_k) \quad \leq \quad (1 - 2\sigma c\lambda_k)\|f(x_k)\|^2.$$

Taking $\alpha := 2\sigma c$, we get $0 < \alpha < 2\sigma$, and (10) follows.

The steepest descent algorithm could be applied to find solutions to nonlinear equations of type (9) whenever the function f has a computable Jacobian. The main advantage of Algorithm 1 relative to the steepest descent method is that no derivative information is needed. On the other hand, note that one cannot assure in general that the steepest descent method will converge to a zero of the function f, but to a critical point of $\|f(\cdot)\|^2$ (for more details, see e.g. [5, Chapter 11]). This is not a concern under strong duplomonotonicity for $\sigma > 0$: in this case, any critical point of $\|f(\cdot)\|^2$ will be a zero of f. Indeed, otherwise one would have $\nabla \left(\|f(\cdot)\|^2\right)(\tilde{x}) = 0$ and $f(\tilde{x}) \neq 0$ for some $\tilde{x} \in \mathbb{R}^m$. Then

$$0 = \nabla \left(\|f(\cdot)\|^2\right)(\tilde{x}) = 2\nabla f(\tilde{x}) f(\tilde{x}),$$

whence, by Proposition 2.8(ii),

$$0 = f(\tilde{x})^T \nabla f(\tilde{x}) f(\tilde{x}) \geq \sigma \|f(\tilde{x})\|^2 > 0,$$

which is a contradiction.

If f is Lipschitz continuous with a known constant $\ell > 0$ and is also strongly duplomonotone for $\sigma > 0$ with constant $\bar{\tau} > 0$, then, as a direct consequence of the characterization in Proposition 2.5, we get

$$\|f(x - \tau f(x))\|^2 \leq (1 - 2\sigma\tau + \ell^2\tau^2)\|f(x)\|^2, \tag{11}$$

for all $x \in \mathbb{R}^m$ and all $0 \leq \tau \leq \bar{\tau}$. The right-hand side of (11) attains its minimum (with respect to $\tau \in [0, \bar{\tau}]$) at $\tau^\star := \min\left\{\sigma/\ell^2, \bar{\tau}\right\}$. Thus, if $\sigma/\ell^2 \leq \bar{\tau}$, we have

$$\|f(x - \tau^\star f(x))\|^2 \leq \left(1 - \frac{\sigma^2}{\ell^2}\right)\|f(x)\|^2. \tag{12}$$

This makes us consider the following variation of Algorithm 1, where the step size is chosen constant.

Algorithm 2: Constant step length $\min\left\{\sigma/\ell^2, \bar{\tau}\right\}$

Fix a precision $\varepsilon > 0$. Choose any $x_0 \in \mathbb{R}^m$, and set $k := 0$ and $\lambda := \min\left\{\sigma/\ell^2, \bar{\tau}\right\}$.
while $\|f(x_k)\| > \varepsilon$ **do**
  $x_{k+1} := x_k - \lambda f(x_k)$;
  $k := k + 1$;
end

As a direct consequence of (12) we have that Algorithm 2 is (globally) linearly convergent to a zero of f, and moreover, the Lipschitz assumption can be relaxed as follows.

Theorem 3.1. *Let $f : \mathbb{R}^m \to \mathbb{R}^m$ be strongly duplomonotone for $\sigma > 0$ with constant $\bar{\tau} > 0$. Let $x_0 \in \mathbb{R}^m$ be an initial point, and assume there exists some constant $\ell > 0$ such that*

$$\|f(x - \tau f(x)) - f(x)\| \leq \ell\tau\|f(x)\| \quad \text{for all } x \in L(x_0) \text{ and all } 0 \leq \tau \leq \bar{\tau}, \tag{13}$$

where $L(x_0)$ *is the* lower level set *defined by*

$$L(x_0) := \{x \in \mathbb{R}^m : \|f(x)\| \le \|f(x_0)\|\}. \tag{14}$$

Set $\lambda := \min\{\sigma/\ell^2, \bar\tau\}$. *Then the iteration* $x_{k+1} := x_k - \lambda f(x_k)$ *satisfies*

$$\|f(x_{k+1})\| \le \sqrt{1 - 2\sigma\lambda + \ell^2\lambda^2}\,\|f(x_k)\|;$$

whence, $f(x_k)$ *is linearly convergent to zero. Thus, if* f *is continuous, any accumulation point of the sequence* x_k *is a zero of* f.

Proof. It follows from the argumentation above. □

Even when f is known to be Lipschitz continuous, its Lipschitz constant might not be easy to compute. The next result shows that in this case Algorithm 1 can be used, and the step size λ_k can always be found by a *backtracking* technique where λ_k is bounded away from zero and the algorithm is linearly convergent. We denote by $\lceil\cdot\rceil$ the *ceiling function*, i.e., the smallest following integer to a given number.

Theorem 3.2. *Let* $f : \mathbb{R}^m \to \mathbb{R}^m$ *be strongly duplomonotone for* $\sigma > 0$ *with constant* $\bar\tau > 0$. *Let* $x_0 \in \mathbb{R}^m$ *be an initial point, and assume that there is a positive constant* ℓ *such that* (13) *holds. Then, for all* $0 < \alpha < 2\sigma$ *and all* $0 < \beta < 1$, *Algorithm 1 generates a sequence* x_k *such that* $f(x_k)$ *is linearly convergent to zero with rate* $\sqrt{1-\beta^p}$, *where*

$$p := \left\lceil \frac{1}{\log\beta}\left(\log\alpha + \min\{\log(\bar\tau), \log(2\sigma - \alpha) - 2\log\ell\}\right)\right\rceil. \tag{15}$$

Thus, if f *is continuous, any accumulation point of the sequence* x_k *is a zero of* f.

Proof. Let $x \in L(x_0)$. We will prove that the step size $(1/\alpha)\beta^p$ with p as in (15) always satisfies (10), i.e., that we have

$$\|f(x - (1/\alpha)\beta^p f(x))\|^2 \le (1 - \beta^p)\|f(x)\|^2. \tag{16}$$

Proposition 2.5 gives us

$$\|f(x - \tau f(x))\|^2 \le (1 - 2\sigma\tau)\|f(x)\|^2 + \|f(x - \tau f(x)) - f(x)\|^2, \tag{17}$$

for all $\tau \in [0, \bar\tau]$. Take p as in (15), that is,

$$p := \left\lceil \max\left\{\frac{\log(\alpha\bar\tau)}{\log\beta}, \frac{\log\alpha + \log(2\sigma - \alpha) - 2\log\ell}{\log\beta}\right\}\right\rceil.$$

Then (13) holds for all $0 < \tau \le (1/\alpha)\beta^p$. This, together with (17), implies that

$$\|f(x - \tau f(x))\|^2 \le (1 - 2\sigma\tau + \ell^2\tau^2)\|f(x)\|^2 \quad \text{whenever } 0 < \tau \le \frac{1}{\alpha}\beta^p.$$

Moreover, we have that $1 - 2\sigma\tau + \ell^2\tau^2 \le 1 - \alpha\tau$ if and only if $\tau \le (2\sigma - \alpha)/\ell^2$. The definition of p implies that $(1/\alpha)\beta^p \le (2\sigma - \alpha)/\ell^2$; hence,

$$\|f(x - \tau f(x))\|^2 \le (1 - \alpha\tau)\|f(x)\|^2 \quad \text{whenever } 0 < \tau \le \frac{1}{\alpha}\beta^p,$$

which implies (16). Therefore, given a point x_k generated by Algorithm 1, the integer p_k can always be found and it satisfies $p_k \le p$. Thus, $\lambda_k = (1/\alpha)\beta^{p_k} \ge (1/\alpha)\beta^p$, and we have

$$\|f(x_{k+1})\|^2 = \|f(x_k - \lambda_k f(x_k)\|^2 \le (1 - \alpha\lambda_k)\|f(x_k)\|^2 \le (1 - \beta^p)\|f(x_k)\|^2,$$

which in particular yields $x_{k+1} \in L(x_0)$, and the claims in the statement follow. □

Remark 3.3. Notice that: (i) the constant ℓ in (13) does not need to be known in order to use Algorithm 1, but is involved in the rate of convergence, and (ii) Lipschitz continuity of the function f on $L(x_0)$ implies (13).

Even if f is known (or conjectured) to be both Lipschitz continuous and strongly duplomonotone for $\sigma > 0$, in practical situations the values of both the Lipschitz constant and σ might be unknown or difficult to compute. The following modification of Algorithm 1 permits finding an adequate step size by a double backtracking technique, where an additional backtracking is performed in order to find an appropriate value of the parameter α in (10) such that $\alpha < 2\sigma$.

Algorithm 3: Double backtracking (with $\sigma > 0$ unknown)

Fix a precision $\varepsilon > 0$, and a minimum and a maximum step size $0 < \lambda_{\min} < \lambda_{\max}$.
Choose any $x_0 \in \mathbb{R}^m$, $0 < \alpha < \lambda_{\max}^{-1}$, $0 < \beta < 1$, and set $k := 0$.
while $\|f(x_k)\| > \varepsilon$ **do**
  $\lambda_k := \lambda_{\max}$;
  while $\|f(x_k - \lambda_k f(x_k))\|^2 > (1 - \alpha\lambda_k)\|f(x_k)\|^2$ **do**
    $\lambda_k := \beta\lambda_k$;
    if $\lambda_k < \lambda_{\min}$ **then**
      $\alpha := \alpha\beta$;
      $\lambda_k := \lambda_{\max}$;
    end
  end
  $x_{k+1} := x_k - \lambda_k f(x_k)$;
  $k := k + 1$;
end

Theorem 3.4. *Let $f : \mathbb{R}^m \to \mathbb{R}^m$ be strongly duplomonotone for $\sigma > 0$ with constant $\bar{\tau} > 0$. Let $x_0 \in \mathbb{R}^m$ be an initial point, and assume that there exists some positive constants ℓ such that (13) holds. Then, for all positive constants $\lambda_{\min}$ and $\lambda_{\max}$ such that there exists some integer q with $\lambda_{\min} \leq \beta^q \lambda_{\max} < \min\left\{2\sigma/\ell^2, \bar{\tau}\right\}$, Algorithm 3 generates a sequence x_k such that $f(x_k)$ is linearly convergent to zero with rate $\sqrt{1 - \alpha\beta^{p+q}\lambda_{\max}}$, where*

$$p := \left\lceil \frac{\log(2\sigma - \ell^2\beta^q\lambda_{\max}) - \log(\alpha)}{\log(\beta)} \right\rceil. \tag{18}$$

Thus, if f is continuous, any accumulation point of the sequence x_k is a zero of f.

Proof. Denote by α_0 the initial value of α in Algorithm 3. Proposition 2.5 together with (13) gives us

$$\|f(x - \tau f(x))\|^2 \leq (1 - 2\sigma\tau + \ell^2\tau^2)\|f(x)\|^2 \quad \text{whenever } x \in L(x_0), 0 < \tau \leq \bar{\tau}.$$

Further, we have that $1 - 2\sigma\tau + \ell^2\tau^2 \leq 1 - \alpha_0\beta^p\tau$ with $0 < \tau \leq \bar{\tau}$ if and only if $0 < \tau \leq \min\left\{(2\sigma - \alpha_0\beta^p)/\ell^2, \bar{\tau}\right\}$. By assumption, there exists some positive integer q such that $\lambda_{\min} \leq \beta^q\lambda_{\max} < \min\left\{2\sigma/\ell^2, \bar{\tau}\right\}$. By the definition of p in (18), we have $\beta^q\lambda_{\max} \leq (2\sigma - \alpha_0\beta^p)/\ell^2$. Hence,

$$\lambda_{\min} \leq \beta^q\lambda_{\max} \leq \min\left\{\frac{2\sigma - \alpha_0\beta^p}{\ell^2}, \bar{\tau}\right\}.$$

Thus, for all $x \in L(x_0)$, we have

$$\|f(x - (\beta^q\lambda_{\max})f(x))\|^2 \leq (1 - \alpha_0\beta^{p+q}\lambda_{\max})\|f(x)\|^2. \tag{19}$$

Finally, observe that there is some positive integer s such that $\beta^s\lambda_{\max} < \lambda_{\min}$. Therefore, given x_k, a new point x_{k+1} is guaranteed to be found in a finite number of steps of Algorithm 3, because the double backtracking loop can only be executed a maximum of $sp + q$ times (after a maximum of sp iterations the value of α will be equal to $\alpha\beta^p$, after which, a maximum of q iterations will be

11

enough to find an appropriate step size λ_k). Thus, we have $\alpha\lambda_k \geq \alpha_0\beta^{p+q}\lambda_{\max}$. Consequently, by the acceptance criteria of the step size in Algorithm 3, we have

$$\|f(x_{k+1})\|^2 \leq (1 - \alpha\lambda_k)\|f(x_k)\|^2 \leq (1 - \alpha_0\beta^{p+q}\lambda_{\max})\|f(x)\|^2,$$

and the claims follow. $\qquad\square$

Remark 3.5. (i) The condition $\lambda_{\min} \leq \beta^q\lambda_{\max} < \min\{2\sigma/\ell^2, \bar{\tau}\}$ in Theorem 3.4 is needed to avoid the possibility of an infinite loop in an iteration of the algorithm. Nevertheless, we believe this condition should not be too difficult to guarantee in practice, as it basically requires that $\lambda_{\min}$ is not "too big" and β is not "too small".
(ii) Certainly, the constant β used for updating α can be chosen different from the constant β used for updating λ_k, and Theorem 3.4 would remain valid with slight changes. Nonetheless, we have decided to use the same constant to ease the notation and the analysis.
(iii) In Algorithm 3, the constant α is required to be smaller than $\lambda_{\max}^{-1}$ to avoid unnecessary iterations (otherwise, the initial step $\lambda_k = \lambda_{\max}$ would always be too big, since $1 - \alpha\lambda_k$ would be negative).

Acknowledgements

References

[1] Heinz H. Bauschke and Patrick L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces.* Springer, April 2011.

[2] Alberto Cambini and Laura Martein. *Generalized Convexity and Optimization: Theory and Applications.* Springer, October 2008.

[3] R. M. T. Fleming and I. Thiele. Mass conserved elementary kinetics is sufficient for the existence of a non-equilibrium steady state concentration. *Journal of Theoretical Biology*, 314:173–181, 2012.

[4] Nicolas Hadjisavvas, Sándor Komlósi, and Siegfried S. Schaible. *Handbook of Generalized Convexity and Generalized Monotonicity.* Springer, 2005.

[5] Jorge Nocedal and Stephen J. Wright. *Numerical Optimization.* Springer Series in Operations Research and Financial Engineering. Springer, New York, second edition, 2006.

[6] R. Tyrrell Rockafellar and Roger J.-B. Wets. *Variational Analysis*, volume 317 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1998.