Chapter 30

Complex images

ap,complex

s,complex,intro

s,complex,min

s,complex,frechet

Contents

30.1 Introduction (s,complex,intro)	30.1
30.2 Minimization over complex vectors (s,complex,min)	30.1
30.2.1 Fréchet differentiability (s,complex,frechet)	30.1
30.2.2 Euclidean norm on $\mathbb{C}^{n_{p}}$ (s,complex,min,norm2)	30.2
30.2.3 Steepest ascent direction for functionals on $\mathbb{C}^{n_{p}}$ (s,complex,min,steep)	30.2
30.2.4 Holomorphic functions (s,complex,holo)	30.3
30.2.5 Quadratic case (s,complex,quad)	30.4
30.2.6 Nonlinear WLS cost functions (s,complex,min,nwls)	30.4
30.2.7 WLS cost functions (s,complex,min,wls)	30.5
30.2.7.1 Step sizes (s,complex,min,step,r)	30.5
30.2.7.2 Complex step sizes? (s,complex,min,step,c)	30.6
30.3 Real constraints (s,complex,real)	30.6
30.3.1 Bound analysis	30.7
30.4 Taylor series (s,complex,taylor)	30.8
30.5 Convex functions of complex arguments (s,complex,convex)	30.9
30.6 Problems (s,complex,prob)	30.11
30.7 Bibliography	30.11

30.1 Introduction (s, complex, intro)

For many of the applications considered in this book, all quantities are real. This appendix addresses special issues that arise when the image x, the data y, and/or the system matrix A have complex elements, such as in MRI. This subject can be subtle [1–6].

30.2 Minimization over complex vectors (s,complex,min)

Often we want to apply gradient-based optimization methods to a cost function $\Psi(x)$ where x has complex elements. However, many of the functionals of interest do not have a **derivative** when viewed in \mathbb{C}^{n_p} . This complication is related to the fact that $|z|^2$ for $z \in \mathbb{C}$ is differentiable only at z = 0 [7, p. 31]. This section addresses such issues in the context of optimization.

30.2.1 Fréchet differentiability (s,complex,frechet)

We say $g : \mathbb{C}^n \to \mathbb{C}^m$ is Fréchet differentiable at $x \in \mathbb{C}^n$ if there exists a bounded linear operator $G_x : \mathbb{C}^n \to \mathbb{C}^m$ such that

$$\lim_{h \to 0} \frac{\|g(x+h) - g(x) - G_x h\|_2}{\|h\|_2} = 0.$$
(30.2.1)

(The limit must not depend on *how* h approaches 0, *i.e.*, we must reach the same limit for *any* sequence $\{h^{(n)}\}$ in \mathbb{C}^n that converges to 0.) The norm in the numerator of (30.2.1) is for \mathbb{C}^m whereas the one in the denominator is for \mathbb{C}^n . The Fréchet differential of g at x, if it exists, is unique [8, p. 172]. Unfortunately, often it does not exist for cost functions of interest in image reconstruction, as shown in the next example.

30.2.2 Euclidean norm on $\mathbb{C}^{n_{p}}$ (s,complex,min,norm2)

Consider the simple Euclidean norm functional $\Psi: \mathbb{C}^{n_{p}} \to \mathbb{R}$ defined for $x \in \mathbb{C}^{n_{p}}$ by

$$\Psi(oldsymbol{x}) = rac{1}{2} \left\|oldsymbol{x}
ight\|^2 = rac{1}{2} \left\langleoldsymbol{x}, \,oldsymbol{x}
ight
angle = rac{1}{2} \sum_{j=1}^{n_{
m p}} \left|x_j
ight|^2.$$

For $\boldsymbol{h} \in \mathbb{C}^{n_{\mathrm{P}}}$ we have

$$\Psi(\boldsymbol{x}+\boldsymbol{h}) - \Psi(\boldsymbol{x}) = \frac{1}{2} \langle \boldsymbol{x}, \boldsymbol{h} \rangle + \frac{1}{2} \langle \boldsymbol{h}, \boldsymbol{x} \rangle + \frac{1}{2} \|\boldsymbol{h}\|^2 = \operatorname{real}\{\langle \boldsymbol{x}, \boldsymbol{h} \rangle\} + \frac{1}{2} \|\boldsymbol{h}\|^2,$$

which is *not* linear in h (even ignoring the ||h|| term), because $\langle x, \alpha h \rangle = \alpha^* \langle x, h \rangle \neq \alpha \langle x, h \rangle$ when α is complex. So the functional Ψ does not have a Fréchet differential [8, p. 172] when we view $\Psi : \mathbb{C}^{n_p} \to \mathbb{R}$.

However, if we separate the real and imaginary parts: $\boldsymbol{x} = \boldsymbol{a} + i\boldsymbol{b}$, where $i = \sqrt{-1}$, then we can instead consider the function $\psi(\boldsymbol{a}, \boldsymbol{b}) \triangleq \Psi(\boldsymbol{a} + i\boldsymbol{b}) = \frac{1}{2} \|\boldsymbol{a}\|^2 + \frac{1}{2} \|\boldsymbol{b}\|^2$. Clearly $\nabla_{\boldsymbol{a}}\psi(\boldsymbol{a}, \boldsymbol{b}) = \boldsymbol{a}$ and $\nabla_{\boldsymbol{b}}\psi(\boldsymbol{a}, \boldsymbol{b}) = \boldsymbol{b}$, so

$$\nabla_{\boldsymbol{a}}\psi + \imath\nabla_{\boldsymbol{b}}\psi = \boldsymbol{a} + \imath\boldsymbol{b} = \boldsymbol{x}.$$

So if we somewhat abuse notation and write

$$abla \Psi(\boldsymbol{x}) =
abla rac{1}{2} \left\| \boldsymbol{x} \right\|^2 = \boldsymbol{x},$$
(30.2.2)

then we must remember that this expression does not satisfy the usual definition of a derivative, *i.e.*, the limit (30.2.1) does not exist, except at x = 0, because in general the limit depends on how $h \in \mathbb{C}^{n_{p}}$ approaches zero. (See [9] for discussion of Wirtinger calculus for differentiation and [10] for its application to phase retrieval.) The next section provides a useful interpretation and generalization of (30.2.2).

30.2.3 Steepest ascent direction for functionals on $\mathbb{C}^{n_{p}}$ (s,complex,min,steep)

Having seen that even the simple Euclidean norm functional involves subtleties, we need a systematic way to work with more general functionals $\Psi : \mathbb{C}^{n_{\mathrm{P}}} \to \mathbb{R}$. Fortunately, for the purposes of optimization algorithms, Fréchet derivatives are not essential. To apply a gradient-based minimization method to suitably smooth cost functions, we need only to be able to identify a **descent direction**. Our starting point for that is the "**direction of steepest ascent**," *i.e.*, the vector $s \in \mathbb{C}^{n_{\mathrm{P}}}$ for which $\Psi(x + \alpha s)$ increases most rapidly for small, *real*, positive values of α . We define such directions using **directional derivatives**.

Definition 30.2.1 A vector $s_* = s_*(x) \in \mathbb{C}^{n_p}$ is a direction of steepest ascent of $\Psi : \mathbb{C}^{n_p} \to \mathbb{R}$ at $x \in \mathbb{C}^{n_p}$ if

$$\lim_{\alpha \searrow 0^+} \frac{\Psi(\boldsymbol{x} + \alpha \, \boldsymbol{s}_*) - \Psi(\boldsymbol{x})}{\alpha} \ge \lim_{\alpha \searrow 0^+} \frac{\Psi(\boldsymbol{x} + \alpha \, \boldsymbol{s}) - \Psi(\boldsymbol{x})}{\alpha}, \ \forall \left\{ \boldsymbol{s} \in \mathbb{C}^{n_{\mathrm{p}}} : \|\boldsymbol{s}\| \le \|\boldsymbol{s}_*\| \right\}, \tag{30.2.3}$$

where α is real.

To analyze (30.2.3), define $f : \mathbb{R} \to \mathbb{R}$ by

$$f(\alpha) \triangleq f(\alpha; \boldsymbol{s}) = \Psi(\boldsymbol{x} + \alpha \, \boldsymbol{s}),$$

for which

$$\dot{f}(0) = \lim_{\alpha \searrow 0^+} \frac{f(\alpha) - f(0)}{\alpha} = \lim_{\alpha \searrow 0^+} \frac{\Psi(\boldsymbol{x} + \alpha \, \boldsymbol{s}) - \Psi(\boldsymbol{x})}{\alpha}$$

As in §30.2.2, define

$$\psi(\boldsymbol{a}, \boldsymbol{b}) \triangleq \Psi(\boldsymbol{a} + \imath \boldsymbol{b})$$

Then by the usual **chain rule** for differentiation:

$$\dot{f}(0) = \lim_{\alpha \searrow 0^+} \frac{\psi(\boldsymbol{x}_{\mathrm{R}} + \alpha \, \boldsymbol{s}_{\mathrm{R}}, \boldsymbol{x}_{\mathrm{I}} + \alpha \, \boldsymbol{s}_{\mathrm{I}}) - \psi(\boldsymbol{x}_{\mathrm{R}}, \boldsymbol{x}_{\mathrm{I}})}{\alpha} = \boldsymbol{s}_{\mathrm{R}}' \, \nabla_{\boldsymbol{a}} \psi(\boldsymbol{x}_{\mathrm{R}}, \boldsymbol{x}_{\mathrm{I}}) + \boldsymbol{s}_{\mathrm{I}}' \, \nabla_{\boldsymbol{b}} \psi(\boldsymbol{x}_{\mathrm{R}}, \boldsymbol{x}_{\mathrm{I}})$$

$$= \operatorname{real}\{\boldsymbol{s}' \, \boldsymbol{g}(\boldsymbol{x})\} = \operatorname{real}\{\langle \boldsymbol{s}, \, \boldsymbol{g}(\boldsymbol{x}) \rangle\},$$

$$(30.2.4)^{\text{e, complex, min, f}}$$

assuming that ψ is differentiable in both arguments a and b, where we define the following complex function:

$$g(\boldsymbol{x}) \triangleq \left. \nabla_{\boldsymbol{a}} \psi(\boldsymbol{a}, \boldsymbol{b}) + \imath \nabla_{\boldsymbol{b}} \psi(\boldsymbol{a}, \boldsymbol{b}) \right|_{\boldsymbol{a}=\boldsymbol{x}_{\mathrm{R}}, \ \boldsymbol{b}=\boldsymbol{x}_{\mathrm{I}}}.$$
(30.2.5)

The expression (30.2.4) for $\dot{f}(0)$ is useful for line search operations.

By the Schwarz inequality (28.4.2):

$$ext{real}\{\langle m{s},\,m{g}
angle\} \leq |\langle m{s},\,m{g}
angle| \leq \|m{s}\|\,\|m{g}\|\,,$$

with equality if g = 0 or $s = \beta g$ for $\beta \in [0, \infty)$. Thus $s_*(x) = g(x)$ is a direction of steepest ascent. To summarize, for the purposes of gradient-based optimization, the following "definition" works:

$$\nabla \Psi(\boldsymbol{x}) \triangleq \nabla_{\boldsymbol{a}} \psi(\boldsymbol{a}, \boldsymbol{b}) + i \nabla_{\boldsymbol{b}} \psi(\boldsymbol{a}, \boldsymbol{b}) \Big|_{\boldsymbol{a} = \boldsymbol{x}_{\mathrm{R}}, \ \boldsymbol{b} = \boldsymbol{x}_{\mathrm{I}}}.$$
(30.2.6)

Throughout this *book* we use "**gradient**" to mean such a direction of steepest ascent (rather than a Fréchet differential) for real-valued functionals of complex arguments. It is also called the **conjugate cogradient** [11] [12]. We call such functionals **differentiable** when the gradients on the right side of (30.2.6) exist.

Although the limit (30.2.1) often does not exist for cost functions of interest on $\mathbb{C}^{n_{p}}$, if $\psi(a, b)$ is differentiable in both arguments (*i.e.*, on $\mathbb{R}^{n_{p}} \times \mathbb{R}^{n_{p}}$) then we do have the following limit:

$$\lim_{\boldsymbol{h}\to 0} \frac{\Psi(\boldsymbol{x}+\boldsymbol{h}) - \Psi(\boldsymbol{x}) - \operatorname{real}\{\langle \nabla \Psi(\boldsymbol{x}), \, \boldsymbol{h} \rangle\}}{\|\boldsymbol{h}\|} = 0,$$

i.e., we have the following property that generalizes the 1st-order Taylor series (29.8.3):

$$\Psi(\boldsymbol{x} + \boldsymbol{h}) = \Psi(\boldsymbol{x}) + \operatorname{real}\{\langle \nabla \Psi(\boldsymbol{x}), \, \boldsymbol{h} \rangle\} + o(\|\boldsymbol{h}\|)$$

$$= \Psi(\boldsymbol{x}) + \operatorname{real}\left\{\int_{0}^{1} \langle \nabla \Psi(\boldsymbol{x} + \tau \boldsymbol{h}), \, \boldsymbol{h} \rangle \, \mathrm{d}\tau\right\}.$$
(30.2.8)

This property suffices for analyzing most optimization problems.

Example 30.2.2 For the cost function $\Psi(\boldsymbol{x}) = \operatorname{real}\{\langle \boldsymbol{b}, \boldsymbol{x} \rangle\} = \langle \boldsymbol{b}_{\mathrm{R}}, \boldsymbol{x}_{\mathrm{R}} \rangle + \langle \boldsymbol{b}_{\mathrm{I}}, \boldsymbol{x}_{\mathrm{I}} \rangle$, the gradient is $\nabla \Psi = \boldsymbol{b}$.

30.2.4 Holomorphic functions (s,complex,holo)

To proceed we will need the following definition.

Definition 30.2.3 A function $g : \mathbb{C}^n \to \mathbb{C}^m$ is called **holomorphic** (often called **analytic**) if it is Fréchet differentiable per (30.2.1) on \mathbb{C}^n .

For our purposes, the key property of holomorphic functions is the following. If $g : \mathbb{C}^n \to \mathbb{C}^m$ is holomorphic, then writing g(a + ib) = u(a, b) + i v(a, b), its components u and v satisfy the Cauchy-Riemann equations:

$$\frac{\partial}{\partial a_j}u_i(\boldsymbol{a},\boldsymbol{b}) = \frac{\partial}{\partial b_j}v_i(\boldsymbol{a},\boldsymbol{b}), \qquad \frac{\partial}{\partial b_j}u_i(\boldsymbol{a},\boldsymbol{b}) = -\frac{\partial}{\partial a_j}v_i(\boldsymbol{a},\boldsymbol{b}), \quad i = 1,\dots,m, \quad j = 1,\dots,n.$$
(30.2.9)

Put another way:

$$abla_{\boldsymbol{a}} \boldsymbol{u}(\boldsymbol{a}, \boldsymbol{b}) =
abla_{\boldsymbol{b}} \boldsymbol{v}(\boldsymbol{a}, \boldsymbol{b}), \qquad -
abla_{\boldsymbol{b}} \boldsymbol{u}(\boldsymbol{a}, \boldsymbol{b}) =
abla_{\boldsymbol{a}} \boldsymbol{v}(\boldsymbol{a}, \boldsymbol{b})$$

Example 30.2.4 The function $g(x) \triangleq Ax$ for $A \in \mathbb{C}^{n_d \times n_p}$ is holomorphic, because $u(a, b) = A_R a - A_I b$, $v(a, b) = A_I a + A_R b$, $\nabla_a u(a, b) = \nabla_b v(a, b) = A_R$, and $-\nabla_b u(a, b) = \nabla_a v(a, b) = A_I$.

Example 30.2.5 Consider the function $g(\mathbf{x}) \triangleq \alpha \mathbf{x}_R + \beta \mathbf{x}_I$, corresponding to the gradient (30.2.6) of a regularizer (2.9.1) where $\mathsf{R}(\mathbf{x}) = \frac{\alpha}{2} \|\mathbf{x}_R\|_2^2 + \frac{\beta}{2} \|\mathbf{x}_I\|_2^2$, motivated by separately regularizing the real and imaginary parts with $\alpha \neq \beta$ [13–16]. In this case $\mathbf{u}(\mathbf{a}, \mathbf{b}) = \alpha \mathbf{a}, \mathbf{v}(\mathbf{a}, \mathbf{b}) = \beta \mathbf{b}, \nabla_{\mathbf{a}} \mathbf{u}(\mathbf{a}, \mathbf{b}) = \alpha \mathbf{I} \neq \nabla_{\mathbf{b}} \mathbf{v}(\mathbf{a}, \mathbf{b}) = \beta \mathbf{I}$, so $g(\mathbf{x})$ is not holomorphic.

For reconstructing complex images, usually the cost functions of interest are not (Fréchet) differentiable, but fortunately often the "gradients," as defined by (30.2.6), are holomorphic. Thankfully, Example 30.2.5 is the exception, not the rule, but one must be aware of such exceptions.

30.2.5 Quadratic case (s,complex,quad)

Consider the quadratic cost function $\Psi: \mathbb{C}^n \to \mathbb{R}$ defined by

$$\Psi(\boldsymbol{x}) = \frac{1}{2} \operatorname{real}\{\boldsymbol{x}' \boldsymbol{M} \boldsymbol{x}\}$$
(30.2.10)

for a general complex matrix $M \in \mathbb{C}^{n \times n}$. Write M = U + iV where $U, V \in \mathbb{R}^{n \times n}$, then

$$\psi(\boldsymbol{a}, \boldsymbol{b}) \triangleq \Psi(\boldsymbol{a} + \imath \boldsymbol{b}) = \frac{1}{2} \operatorname{real}\{(\boldsymbol{a} + \imath \boldsymbol{b})'(\boldsymbol{U} + \imath \boldsymbol{V})(\boldsymbol{a} + \imath \boldsymbol{b})\}$$
$$= \frac{1}{2} (\boldsymbol{a}'\boldsymbol{U}\boldsymbol{a} - \boldsymbol{a}'\boldsymbol{V}\boldsymbol{b} + \boldsymbol{b}'\boldsymbol{V}\boldsymbol{a} + \boldsymbol{b}'\boldsymbol{U}\boldsymbol{b}).$$

Thus using (29.8.2) the gradient components are

$$u = \nabla_a \psi = \frac{1}{2} (U + U')a - \frac{1}{2} (V - V')b = Sa - Tb$$
$$v = \nabla_b \psi = \frac{1}{2} (U + U')b + \frac{1}{2} (V - V')b = Sb + Ta,$$

where we define the following symmetric real matrix S and anti-symmetric real matrix T:

$$oldsymbol{S} riangleq rac{1}{2}(oldsymbol{U}+oldsymbol{U}') \ oldsymbol{T} riangleq rac{1}{2}(oldsymbol{V}-oldsymbol{V}').$$

Combining yields the gradient expression

$$abla \Psi = oldsymbol{u} + \imath oldsymbol{v} = (oldsymbol{S} + \imath oldsymbol{T})oldsymbol{x} = rac{1}{2}(oldsymbol{M} + oldsymbol{M}')oldsymbol{x}.$$

In most (if not all) cases of interest, M is Hermitian symmetric, *e.g.*, for the quadratic cost function $\Psi(\mathbf{x}) = \frac{1}{2} \|A\mathbf{x}\|^2$ we have M = A'A. In such cases the "real" operator in (30.2.10) is unnecessary and the gradient expression simplifies:

$$abla \Psi =
abla rac{1}{2} oldsymbol{x}' oldsymbol{M} oldsymbol{x} = oldsymbol{M} oldsymbol{x}$$

Note that

$$\nabla_{\boldsymbol{b}}\boldsymbol{u} = \nabla_{\boldsymbol{b}}\nabla_{\boldsymbol{a}}\psi = -\boldsymbol{I}$$
$$\nabla_{\boldsymbol{a}}\boldsymbol{v} = \nabla_{\boldsymbol{a}}\nabla_{\boldsymbol{b}}\psi = \boldsymbol{T}.$$

Thus the symmetry of second derivatives holds:

$$\nabla_{\boldsymbol{b}} \nabla_{\boldsymbol{a}} \psi = \left(\nabla_{\boldsymbol{a}} \nabla_{\boldsymbol{b}} \psi \right)'.$$

Furthermore,

s.complex.min.nwls

$$abla_{oldsymbol{a}} oldsymbol{u} =
abla_{oldsymbol{b}} oldsymbol{v} = oldsymbol{S} \ -
abla_{oldsymbol{b}} oldsymbol{u} =
abla_{oldsymbol{a}} oldsymbol{v} = oldsymbol{T},$$

so the complex gradient function $\nabla \Psi = u + iv$ is holomorphic.

30.2.6 Nonlinear WLS cost functions (s,complex,min,nwls)

Now specialize §30.2.3 by considering the WLS cost function

$$\Psi(\boldsymbol{x}) = \frac{1}{2} \left(\bar{\boldsymbol{y}}(\boldsymbol{x}) - \boldsymbol{y} \right)' \boldsymbol{W} \left(\bar{\boldsymbol{y}}(\boldsymbol{x}) - \boldsymbol{y} \right), \qquad (30.2.11)$$

where $x \in \mathbb{C}^{n_{p}}$, W is Hermitian positive-semidefinite, $y \in \mathbb{C}^{n_{d}}$ and $\bar{y} : \mathbb{C}^{n_{p}} \to \mathbb{C}^{n_{d}}$ is possibly nonlinear. Defining

$$\boldsymbol{e}(\boldsymbol{x}) \triangleq \boldsymbol{W}^{1/2}\left(\bar{\boldsymbol{y}}(\boldsymbol{x}) - \boldsymbol{y}\right) = \boldsymbol{u}(\boldsymbol{x}_{\mathrm{R}}, \boldsymbol{x}_{\mathrm{I}}) + \imath \, \boldsymbol{v}(\boldsymbol{x}_{\mathrm{R}}, \boldsymbol{x}_{\mathrm{I}}), \qquad (30.2.12)$$

we would like to apply (30.2.6) to the following:

$$\psi(a, b) = \frac{1}{2} \left(u(a, b) + i v(a, b) \right)' \left(u(a, b) + i v(a, b) \right) = \frac{1}{2} \left\| u(a, b) \right\|^2 + \frac{1}{2} \left\| v(a, b) \right\|^2.$$

To find a concise and convenient expression, we assume that e(x) is holomorphic, per Definition 30.2.3. Under this assumption,

$$\begin{pmatrix} \frac{\partial}{\partial a_j} + i \frac{\partial}{\partial b_j} \end{pmatrix} \psi(\boldsymbol{a}, \boldsymbol{b}) = \sum_{i=1}^{n_d} \left(\frac{\partial}{\partial a_j} u_i + i \frac{\partial}{\partial b_j} u_i \right) u_i + \sum_{i=1}^{n_d} \left(\frac{\partial}{\partial a_j} v_i + i \frac{\partial}{\partial b_j} v_i \right) v_i$$

$$= \sum_{i=1}^{n_d} \left(\frac{\partial}{\partial a_j} u_i - i \frac{\partial}{\partial a_j} v_i \right) u_i + \sum_{i=1}^{n_d} \left(\frac{\partial}{\partial a_j} v_i + i \frac{\partial}{\partial a_j} u_i \right) v_i$$

$$= \sum_{i=1}^{n_d} \left(\frac{\partial}{\partial a_j} u_i + i \frac{\partial}{\partial a_j} v_i \right)^* (u_i + i v_i) = \sum_{i=1}^{n_d} g_{ij}^* e_i = [\boldsymbol{G}' \boldsymbol{e}]_i ,$$

where \boldsymbol{G} is the $n_{\mathrm{d}} \times n_{\mathrm{p}}$ matrix having elements

$$g_{ij} \triangleq \frac{\partial}{\partial a_j} u_i + i \frac{\partial}{\partial a_j} v_i.$$

Thus, if $\bar{y}(x)$ is holomorphic, the direction of steepest ascent is

$$\nabla \Psi(\boldsymbol{x}) = \left(\nabla_{\boldsymbol{x}} \, \bar{\boldsymbol{y}}(\boldsymbol{x})\right)' \boldsymbol{W} \left(\bar{\boldsymbol{y}}(\boldsymbol{x}) - \boldsymbol{y}\right), \qquad (30.2.13)$$

where we define $\nabla_x \, \bar{y}(x)$ to be the $n_{\rm d} \times n_{\rm p}$ matrix having elements

$$[\nabla_{\boldsymbol{x}} \, \bar{\boldsymbol{y}}(\boldsymbol{x})]_{ij} = \frac{\partial}{\partial a_j} \operatorname{real}\{\bar{y}_i(a_j + \imath b_j)\} + \imath \frac{\partial}{\partial a_j} \operatorname{imag}\{\bar{y}_i(a_j + \imath b_j)\}.$$
(30.2.14)

This matrix depends only on the derivatives with respect to the real part of x.

See Problem 30.4 for an example related to MRI parameter estimation.

s, complex, min, wis 30.2.7 WLS cost functions (s, complex, min, wls)

A particularly important special case of (30.2.11) is the linear model where $\bar{y}(x) = Ax$, which one can verify is **holomorphic**, for which

$$\Psi(\boldsymbol{x}) = \frac{1}{2} \left(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y} \right)' \boldsymbol{W} \left(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y} \right). \tag{30.2.15}$$

In this case one can check that $\nabla_x \bar{y}(x) = A$, so the direction of steepest ascent (30.2.13) simplifies to

$$\nabla \Psi(\boldsymbol{x}) = \boldsymbol{A}' \boldsymbol{W} \left(\boldsymbol{A} \boldsymbol{x} - \boldsymbol{y} \right). \tag{30.2.16}$$

More generally, if for some $\phi : \mathbb{R} \to \mathbb{R}$ we have

$$\Psi(\boldsymbol{x}) = \phi\bigg(\frac{1}{2}(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y})'\boldsymbol{W}(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y})\bigg),$$

then by similar arguments the corresponding gradient, in the sense of (30.2.6), is:

$$\nabla \Psi(\boldsymbol{x}) = \dot{\phi} \left(\frac{1}{2} (\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y})' \boldsymbol{W} (\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}) \right) \boldsymbol{A}' \boldsymbol{W} (\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}). \tag{30.2.17}^{\text{e, complex, grad, Kx,}}$$

30.2.7.1 Step sizes (s,complex,min,step,r)

s,complex,min,step,r

After choosing a search direction d (such as $-\nabla \Psi(\mathbf{x}^{(n)})$), the steepest descent direction, or alternatively by the conjugate gradients method), often the next problem in a gradient-based algorithm is to find the step size:

$$\alpha_* = \operatorname*{arg\,min}_{\alpha \in [0,\infty)} f(\alpha), \quad f(\alpha) = \Psi(\boldsymbol{x} + \alpha \boldsymbol{d}). \tag{30.2.18}$$

complex.min.nwls.cgrad

For the WLS case in (30.2.15), one can verify:

$$\Psi(\boldsymbol{x} + \alpha \boldsymbol{d}) - \Psi(\boldsymbol{x}) = \frac{1}{2} \alpha^2 (\boldsymbol{A} \boldsymbol{d})' \boldsymbol{W} \boldsymbol{A} \boldsymbol{d} + \alpha \operatorname{real} \{ \boldsymbol{d}' \, \nabla \Psi(\boldsymbol{x}) \} \,.$$

Letting $\boldsymbol{g} = \nabla \Psi(\boldsymbol{x}) = \boldsymbol{A}' \boldsymbol{W} (\boldsymbol{A} \boldsymbol{x} - \boldsymbol{y})$, the minimizing α is

$$\alpha = \frac{-\operatorname{real}\{d'\,g\}}{(Ad)'WAd}.\tag{30.2.19}$$

For non-quadratic problems, one can often find an approximate minimizer of (30.2.18) using Newton's method:

$$\alpha_0 = -\dot{f}(0)/\ddot{f}(0),$$

where $\dot{f}(0) = \operatorname{real}\{d' \, \nabla \Psi(x)\}$ per (30.2.4). Generalizing, one can verify that $\dot{f}(\alpha) = \operatorname{real}\{d' \, \nabla \Psi(x + \alpha d)\}$, so

$$\ddot{f}(0) \approx \frac{\dot{f}(\epsilon) - \dot{f}(0)}{\epsilon} = \frac{1}{\epsilon} \operatorname{real} \{ \boldsymbol{d}' \left[\nabla \Psi(\boldsymbol{x} + \epsilon \boldsymbol{d}) - \nabla \Psi(\boldsymbol{x}) \right] \}.$$

Using this requires choosing ϵ , which is not unlike choosing α in the first place. Indeed the "best" choice for ϵ would be $\epsilon = \alpha_*$ because then $\dot{f}(\alpha_*) = 0$, so the above approximation would yield $\ddot{f}(0) = -\dot{f}(0)/\alpha_*$ and the Newton step size would be exactly α_* . But of course α_* is unknown, so one must use other heuristics to choose ϵ . One should verify that $\Psi(\boldsymbol{x} + \alpha_0 \boldsymbol{d}) < \Psi(\boldsymbol{x})$. (See §11.5.)

30.2.7.2 Complex step sizes? (s,complex,min,step,c)

In (30.2.18) we restricted the step size α to be real, which is the usual practice. As a curiosity, what would happen if we allowed $\alpha \in \mathbb{C}$? Then we would have

$$\Psi(\boldsymbol{x} + \alpha \boldsymbol{d}) - \Psi(\boldsymbol{x}) = \frac{1}{2} \left| \alpha \right|^2 (\boldsymbol{A}\boldsymbol{d})' \boldsymbol{W} \boldsymbol{A} \boldsymbol{d} + \operatorname{real} \{ \alpha^* \boldsymbol{d}' \, \boldsymbol{g} \},$$

where $g(x) = \nabla \Psi(x)$. Let $\gamma = (Ad)' W Ad$, which is real and positive, and define $\beta = d' g / \gamma$. Then

$$\frac{\Psi(\boldsymbol{x} + \alpha \boldsymbol{d}) - \Psi(\boldsymbol{x})}{\gamma} = \frac{1}{2} \left| \alpha \right|^2 + \operatorname{real}\{\alpha^*\beta\} = \frac{1}{2} \left[|\alpha + \beta|^2 - |\beta|^2 \right],$$

so clearly the minimizing $\alpha \in \mathbb{C}$ is

$$\alpha = -\beta = \frac{-d'g}{(Ad)'WAd}.$$
(30.2.20)

For PSD, the search direction is d = -Pg, where P is Hermitian positive definite, so -d'g = g'Pg is real and nonnegative, so the choices (30.2.19) and (30.2.20) for α are identical. But for other methods for choosing the search direction, such as PCG, it seems *plausible* that d'g could be complex. However, see [17, Section 3] for a PCG method where α is always real.

If the minimizing α were complex, then we could write it as $\alpha = |\alpha| e^{i \angle \alpha}$ and define a modified direction vector $\tilde{d} = e^{i \angle \alpha} d$. With respect to this new vector the minimizing α would again be real¹. So if α is complex, it must be due to some "defect" in the direction vector. I do not know if using a complex α would ever be beneficial.

30.3 Real constraints (s,complex,real)

In some applications with complex data y, we know that x should be real, *e.g.*, [18]. (Or essentially equivalently, we may know the phase of each element x_j , so we only want to estimate the magnitude. See also §7.23.) Often in such applications we must minimize a WLS cost function of the form

$$\Psi(\boldsymbol{x}) = \frac{1}{2} \left\| \bar{\boldsymbol{y}}(\boldsymbol{x}) - \boldsymbol{y} \right\|_{\boldsymbol{W}^{1/2}}^2$$

¹Thanks to Angel Pineda for this remark.

over $\boldsymbol{x} \in \mathbb{R}^{n_{\mathrm{p}}}$, where \boldsymbol{W} is diagonal with real elements, $\boldsymbol{y} \in \mathbb{C}^{n_{\mathrm{d}}}$ and $\bar{\boldsymbol{y}} : \mathbb{R}^{n_{\mathrm{p}}} \to \mathbb{C}^{n_{\mathrm{d}}}$. In this case, $\Psi : \mathbb{R}^{n_{\mathrm{p}}} \to \mathbb{R}$, so the gradient of Ψ must be a real vector even though the data is complex. In particular,

$$\begin{split} \Psi(\boldsymbol{x}) &= \frac{1}{2} \left\| \boldsymbol{y} \right\|_{\boldsymbol{W}^{1/2}}^2 - \operatorname{real} \{ \boldsymbol{y}' \boldsymbol{W} \, \bar{\boldsymbol{y}}(\boldsymbol{x}) \} + \frac{1}{2} \left\| \bar{\boldsymbol{y}}(\boldsymbol{x}) \right\|_{\boldsymbol{W}^{1/2}}^2 \\ &\equiv - \boldsymbol{y}_R' \boldsymbol{W} \, \bar{\boldsymbol{y}}_R(\boldsymbol{x}) + \boldsymbol{y}_I' \boldsymbol{W} \, \bar{\boldsymbol{y}}_I(\boldsymbol{x}) + \frac{1}{2} \left\| \bar{\boldsymbol{y}}_R(\boldsymbol{x}) \right\|_{\boldsymbol{W}^{1/2}}^2 + \frac{1}{2} \left\| \bar{\boldsymbol{y}}_I(\boldsymbol{x}) \right\|_{\boldsymbol{W}^{1/2}}^2, \end{split}$$

where $y = y_R + i y_I$. Taking the gradient and collecting like terms, one can verify that:

$$\nabla \Psi(\boldsymbol{x}) = \operatorname{real}\left\{ \left(\bar{\boldsymbol{y}}(\boldsymbol{x}) - \boldsymbol{y} \right)' \boldsymbol{W} \nabla \bar{\boldsymbol{y}}(\boldsymbol{x}) \right\},$$
(30.3.1)

where we define

$$abla ar{m{y}}(m{x}) \triangleq
abla \operatorname{real}\{ar{m{y}}(m{x})\} + i
abla \operatorname{imag}\{ar{m{y}}(m{x})\}$$

Example 30.3.1 For the WLS cost function $\Psi(\mathbf{x}) = \frac{1}{2} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_{\mathbf{W}^{1/2}}^2$, we have $\bar{\mathbf{y}}(\mathbf{x}) = \mathbf{A}\mathbf{x}$. When \mathbf{x} is real, $\nabla \bar{\mathbf{y}} = \mathbf{A}$, and by (30.3.1) the gradient is:

$$\nabla \Psi(\boldsymbol{x}) = \operatorname{real}\{\boldsymbol{A}'\boldsymbol{W}(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y})\} = \operatorname{real}\{\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A}\}\boldsymbol{x} - \operatorname{real}\{\boldsymbol{A}'\boldsymbol{W}\boldsymbol{y}\}.$$
(30.3.2)

So even when A and y are complex, we can use existing methods for multiplication by A and A' to evaluate the gradient with respect to x. Then simply take the real part at the end.

It is clear from (30.3.2) that the Hessian matrix in this case is

$$\nabla^2 \Psi(\boldsymbol{x}) = \operatorname{real}\{\boldsymbol{A}' \boldsymbol{W} \boldsymbol{A}\}. \tag{30.3.3}$$

Example 30.3.2 Continuing Example 30.3.1, it follows from (30.3.2) that the minimizer satisfies real{A'Wy} = real{ $A'WA\hat{x}$ }. Equivalently, because \hat{x} is real:

$$\hat{\boldsymbol{x}} = \left[\operatorname{real}\{\boldsymbol{A}'\boldsymbol{W}\boldsymbol{A}\}\right]^{-1}\operatorname{real}\{\boldsymbol{A}'\boldsymbol{W}\boldsymbol{y}\}.$$
(30.3.4)

(Note that it is not necessary to split A into separate real and imaginary parts as some authors have done [18].) Thus

$$\mathsf{Cov}{\hat{x}} = [\operatorname{real}{A'WA}]^{-1} \mathsf{Cov}{\operatorname{real}{A'Wy}} [\operatorname{real}{A'WA}]^{-1}$$

To simplify, let $\mathbf{y} = \mathbf{y}_{\mathrm{R}} + \mathbf{y}_{\mathrm{I}}$ and assume \mathbf{y}_{R} and \mathbf{y}_{I} are independent with $\mathsf{Cov}\{\mathbf{y}_{\mathrm{R}}\} = \mathsf{Cov}\{\mathbf{y}_{\mathrm{I}}\} = \frac{1}{2}\mathbf{W}^{-1}$, so \mathbf{W} is real valued. Then $\mathsf{Cov}\{\mathbf{y}_{\mathrm{R}}\} = \mathsf{Cov}\{\mathbf{y}_{\mathrm{R}}\} + \mathsf{Cov}\{\mathbf{y}_{\mathrm{R}}\} = \mathbf{W}^{-1}$ and

$$\begin{aligned} \mathsf{Cov}\{\mathrm{real}\{\mathbf{A}'\mathbf{W}\mathbf{y}\}\} &= \mathsf{Cov}\{\mathrm{real}\{\mathbf{A}'\mathbf{W}\}\,\mathbf{y}_{\mathrm{R}} - \mathrm{imag}\{\mathbf{A}'\mathbf{W}\}\,\mathbf{y}_{\mathrm{I}}\} \\ &= \mathrm{real}\{\mathbf{A}'\mathbf{W}\}\,\frac{1}{2}\mathbf{W}^{-1}\left(\mathrm{real}\{\mathbf{A}'\mathbf{W}\}\right)^{T} + \mathrm{imag}\{\mathbf{A}'\mathbf{W}\}\,\frac{1}{2}\mathbf{W}^{-1}\left(\mathrm{imag}\{\mathbf{A}'\mathbf{W}\}\right)^{T} \\ &= \frac{1}{2}\mathbf{A}_{\mathrm{R}}\mathbf{W}\mathbf{A}_{\mathrm{R}}' + \frac{1}{2}\mathbf{A}_{\mathrm{I}}\mathbf{W}\mathbf{A}_{\mathrm{I}}' = \frac{1}{2}\,\mathrm{real}\{\mathbf{A}'\mathbf{W}\}\,.\end{aligned}$$

After choosing a real search direction $d \in \mathbb{R}^{n_{p}}$, *e.g.*, based on the negative of the gradient in (30.3.2), for a linesearch type of method we must find the step $\alpha \in \mathbb{R}$ that minimizes $\Psi(\boldsymbol{x}_{R} + \alpha \boldsymbol{d})$. From the analysis in §30.2, the minimizer is clearly $\alpha = \frac{-d'g}{(Ad)'WAd}$, where $\boldsymbol{g} = \operatorname{real}\{A'W(A\boldsymbol{x}_{R} - \boldsymbol{y})\}$.

So by judicious addition of real statements, existing QPWLS-PCG routines are adapted easily to the case where A and y are complex but x is real.

30.3.1 Bound analysis

Because $x = x_{\rm R} + ix_{\rm I}$, for general complex problems one must estimate both $x_{\rm R}$ and $x_{\rm I}$. On the other hand, if x is known to be real, then we only need to estimate $x_{\rm R}$, *i.e.*, half as many unknown parameters. Having fewer unknowns should reduce the estimator variance. This section uses the **Cramer-Rao lower bound** (CRLB) to quantify this reduction. See also [19].

Because $A = A_{\rm R} + iA_{\rm I}$, we have

$$oldsymbol{A}oldsymbol{x} = (oldsymbol{A}_{ ext{R}} + \imatholdsymbol{A}_{ ext{I}}) = (oldsymbol{A}_{ ext{R}}oldsymbol{x}_{ ext{R}} - oldsymbol{A}_{ ext{I}}oldsymbol{x}_{ ext{I}}) + \imath(oldsymbol{A}_{ ext{I}}oldsymbol{x}_{ ext{R}} + oldsymbol{A}_{ ext{R}}oldsymbol{x}_{ ext{I}})$$

e.complex.real.wls.hess

$$\boldsymbol{A}'\boldsymbol{A} = (\boldsymbol{A}_{\mathrm{R}}' - \imath \boldsymbol{A}_{\mathrm{I}}')(\boldsymbol{A}_{\mathrm{R}} + \imath \boldsymbol{A}_{\mathrm{I}}) = (\boldsymbol{A}_{\mathrm{R}}'\boldsymbol{A}_{\mathrm{R}} + \boldsymbol{A}_{\mathrm{I}}'\boldsymbol{A}_{\mathrm{I}}) + \imath (\boldsymbol{A}_{\mathrm{R}}'\boldsymbol{A}_{\mathrm{I}} - \boldsymbol{A}_{\mathrm{I}}'\boldsymbol{A}_{\mathrm{R}}).$$

Expanding the model (6.1.9) where x is complex:

$$\left[egin{array}{c} m{y}_{ ext{\tiny R}} \ m{y}_{ ext{\tiny I}} \end{array}
ight] = \left[egin{array}{c} m{A}_{ ext{\tiny R}} & -m{A}_{ ext{\tiny I}} \ m{A}_{ ext{\tiny R}} \end{array}
ight] \left[egin{array}{c} m{x}_{ ext{\tiny R}} \ m{x}_{ ext{\tiny I}} \end{array}
ight] + \left[egin{array}{c} m{arepsilon}_{ ext{\tiny R}} \ m{arepsilon}_{ ext{\tiny I}} \end{array}
ight].$$

For white gaussian noise with variance σ^2 , the corresponding Fisher information is

$$\frac{1}{\sigma^2} \begin{bmatrix} \mathbf{A}_{\mathrm{R}}' & \mathbf{A}_{\mathrm{I}}' \\ -\mathbf{A}_{\mathrm{I}}' & \mathbf{A}_{\mathrm{R}}' \end{bmatrix} \begin{bmatrix} \mathbf{A}_{\mathrm{R}} & -\mathbf{A}_{\mathrm{I}} \\ \mathbf{A}_{\mathrm{I}} & \mathbf{A}_{\mathrm{R}} \end{bmatrix} = \frac{1}{\sigma^2} \begin{bmatrix} \mathbf{A}_{\mathrm{R}}' \mathbf{A}_{\mathrm{R}} + \mathbf{A}_{\mathrm{I}}' \mathbf{A}_{\mathrm{I}} & \mathbf{A}_{\mathrm{I}}' \mathbf{A}_{\mathrm{R}} - \mathbf{A}_{\mathrm{R}}' \mathbf{A}_{\mathrm{I}} \\ \mathbf{A}_{\mathrm{I}}' \mathbf{A}_{\mathrm{I}} - \mathbf{A}_{\mathrm{I}}' \mathbf{A}_{\mathrm{R}} & \mathbf{A}_{\mathrm{I}}' \mathbf{A}_{\mathrm{R}} \end{bmatrix} = \frac{1}{\sigma^2} \begin{bmatrix} \mathbf{R} & \mathbf{M}' \\ \mathbf{M} & \mathbf{R} \end{bmatrix} = \frac{1}{\sigma^2} \begin{bmatrix} \mathbf{R} & \mathbf{M}' \\ \mathbf{M} & \mathbf{R} \end{bmatrix}$$

where $\mathbf{R} = \operatorname{real}\{\mathbf{A}'\mathbf{A}\}$ and $\mathbf{M} = \operatorname{imag}\{\mathbf{A}'\mathbf{A}\}$.

Using (28.1.11), the covariance of an unbiased estimate of the real component of x has the following lower bound

$$\mathsf{Cov}\{\hat{x}_{\mathrm{R}}\} \succeq \sigma^2 \left[\mathbf{R} - \mathbf{M}' \mathbf{R}^{-1} \mathbf{M}
ight]^{-1}$$

The model where x is assumed to be real is:

$$\left[egin{array}{c} oldsymbol{y}_{ ext{\tiny R}} \ oldsymbol{y}_{ ext{\tiny I}} \end{array}
ight] = \left[egin{array}{c} oldsymbol{A}_{ ext{\tiny R}} \ oldsymbol{A}_{ ext{\tiny R}} \end{array}
ight] oldsymbol{x}_{ ext{\tiny R}} + \left[egin{array}{c} oldsymbol{arepsilon}_{ ext{\tiny R}} \ oldsymbol{arepsilon}_{ ext{\tiny I}} \end{array}
ight],$$

so the Fisher information is

$$rac{1}{\sigma^2}\left[egin{array}{c} m{A}_{ ext{\tiny R}}^{\prime}m{A}_{ ext{\tiny R}}^{\prime}+m{A}_{ ext{\tiny I}}^{\prime}m{A}_{ ext{\tiny I}}^{\prime}
ight] =rac{1}{\sigma^2}m{R}.$$

In this case the CRLB is

$$\mathsf{Cov}\{\hat{\boldsymbol{x}}_{\scriptscriptstyle \mathrm{R}}\} \succeq \sigma^2 \boldsymbol{R}^{-1}.$$

By Theorem 28.2.4

x, complex, real, A, 2x2

$$\left[\boldsymbol{R}-\boldsymbol{M}'\boldsymbol{R}^{-1}\boldsymbol{M}\right]^{-1}\succeq \boldsymbol{R}^{-1},$$

so as expected the variance is smaller when we enforce the constraint that x is real.

If the columns of A are orthogonal, then M = 0 and the two cases lead to the same covariance of x_{R} . But for nonuniform frequency samples, the columns of A are not orthogonal.

Example 30.3.3 If
$$\mathbf{A} = \begin{bmatrix} 1 & 1 \\ e^{-i\alpha} & e^{-i\beta} \end{bmatrix}$$
, then one can show that
$$\begin{bmatrix} \mathbf{R} - \mathbf{M}' \mathbf{R}^{-1} \mathbf{M} \end{bmatrix}^{-1} = \lambda \mathbf{R}^{-1},$$
(30.3.5)

where $\lambda = \frac{3 + \cos(\alpha - \beta)}{2} \in [1, 2]$. So the covariance without the real constraint can be up to a factor of two larger than the constrained case.

30.4 Taylor series (s,complex,taylor)

This section examines Taylor series in the complex case of a function $\Psi : \mathbb{C}^{n_{p}} \to \mathbb{R}$. Assume that $\psi(a, b) = \Psi(a + ib)$ is twice differentiable on $\mathbb{R}^{n} \times \mathbb{R}^{n}$, and define

$$oldsymbol{u} riangleq
abla_{oldsymbol{a}} \psi \ oldsymbol{v} riangleq
abla_{oldsymbol{b}} \psi \ oldsymbol{v} riangleq
abla_{oldsymbol{b}} \psi \ oldsymbol{a}$$

^

Now expand ψ around (0, 0) use the multivariate 2nd-order Taylor series (29.8.4):

$$\psi(\boldsymbol{a},\boldsymbol{b}) = \psi(\boldsymbol{0},\boldsymbol{0}) : + \begin{bmatrix} \boldsymbol{u}(\boldsymbol{0}) \\ \boldsymbol{v}(\boldsymbol{0}) \end{bmatrix}' \begin{bmatrix} \boldsymbol{a} \\ \boldsymbol{b} \end{bmatrix} + \begin{bmatrix} \boldsymbol{a} \\ \boldsymbol{b} \end{bmatrix}' \int_{0}^{1} (1-\tau) (\nabla^{2}\psi)(\tau \boldsymbol{a},\tau \boldsymbol{b}) \, \mathrm{d}\tau \begin{bmatrix} \boldsymbol{a} \\ \boldsymbol{b} \end{bmatrix}$$
$$= \Psi(\boldsymbol{0}) + \mathrm{real}\{\langle \nabla \Psi(\boldsymbol{0}), \boldsymbol{x} \rangle\} + \begin{bmatrix} \boldsymbol{a} \\ \boldsymbol{b} \end{bmatrix}' \int_{0}^{1} (1-\tau) (\nabla^{2}\psi)(\tau \boldsymbol{a},\tau \boldsymbol{b}) \, \mathrm{d}\tau \begin{bmatrix} \boldsymbol{a} \\ \boldsymbol{b} \end{bmatrix},$$

where $\nabla \Psi = u + iv$, per the convention (30.2.6), and

$$(\nabla^2 \psi)(\boldsymbol{a}, \boldsymbol{b}) = \begin{bmatrix} \nabla_{\boldsymbol{a}}^2 \psi & \nabla_{\boldsymbol{b}} \nabla_{\boldsymbol{a}} \psi \\ \nabla_{\boldsymbol{a}} \nabla_{\boldsymbol{b}} \psi & \nabla_{\boldsymbol{b}}^2 \psi \end{bmatrix} = \begin{bmatrix} \nabla_{\boldsymbol{a}}^2 \psi & \nabla_{\boldsymbol{b}} \boldsymbol{u} \\ \nabla_{\boldsymbol{a}} \boldsymbol{v} & \nabla_{\boldsymbol{b}}^2 \psi \end{bmatrix}.$$

To proceed, we assume that $\nabla \Psi$ is **holomorphic** per Definition 30.2.3, *i.e.*, satisfies

$$abla_{oldsymbol{a}}oldsymbol{u} =
abla_{oldsymbol{b}}oldsymbol{v} riangleq oldsymbol{S} \ -
abla_{oldsymbol{b}}oldsymbol{u} =
abla_{oldsymbol{a}}oldsymbol{v} riangleq oldsymbol{T}.$$

Note that $S = \nabla_a^2 \psi = \nabla_b^2 \psi$ is symmetric, and $T = \nabla_a v = \nabla_a \nabla_b \psi = (\nabla_b \nabla_a \psi)' = (\nabla_b u)' = -T'$ is anti-symmetric.

Expanding the quadratic form:

$$\left[egin{array}{cc} a \\ b \end{array}
ight]' \left[egin{array}{cc}
abla_a^2\psi &
abla_b^au \\
abla_av &
abla_b^2\psi \end{array}
ight] \left[egin{array}{cc} a \\ b \end{array}
ight] = a'Sa - a'Tb + b'Ta + b'Sb.$$

Defining H = S + iT, one can rewrite the quadratic form as x'Hx. Thus we define the (Hermitian) Hessian of Ψ as

$$\nabla^2 \Psi = \mathbf{S} + i\mathbf{T} = \nabla_{\mathbf{a}}^2 \psi + i \nabla_{\mathbf{a}} \nabla_{\mathbf{b}} \psi, \qquad (30.4.1)$$

provided $\nabla \Psi$ is holomorphic. With this definition, we can write the 2nd-order Taylor series as

$$\Psi(\boldsymbol{x}) = \Psi(\boldsymbol{z}) + \operatorname{real}\{\langle \nabla \Psi(\boldsymbol{z}), \, \boldsymbol{x} - \boldsymbol{z} \rangle\} + (\boldsymbol{x} - \boldsymbol{z})' \int_0^1 (1 - \tau) \, \nabla^2 \Psi(\tau \boldsymbol{x} + (1 - \tau) \boldsymbol{z}) \, \mathrm{d}\tau \, (\boldsymbol{x} - \boldsymbol{z}) \,. \tag{30.4.2}$$

Similarly, using the multivariate 1st-order Taylor series (29.8.3):

$$\begin{bmatrix} \boldsymbol{u}(\boldsymbol{a},\boldsymbol{b})\\ \boldsymbol{v}(\boldsymbol{a},\boldsymbol{b}) \end{bmatrix} = \begin{bmatrix} \boldsymbol{u}(\boldsymbol{0},\boldsymbol{0})\\ \boldsymbol{v}(\boldsymbol{0},\boldsymbol{0}) \end{bmatrix} + \int_{0}^{1} (\nabla^{2}\psi)(\tau\boldsymbol{a},\tau\boldsymbol{b}) \,\mathrm{d}\tau \begin{bmatrix} \boldsymbol{a}\\ \boldsymbol{b} \end{bmatrix}$$
$$= \begin{bmatrix} \boldsymbol{u}(\boldsymbol{0},\boldsymbol{0})\\ \boldsymbol{v}(\boldsymbol{0},\boldsymbol{0}) \end{bmatrix} + \int_{0}^{1} \begin{bmatrix} \boldsymbol{S} & -\boldsymbol{T}\\ \boldsymbol{T} & \boldsymbol{S} \end{bmatrix} \,\mathrm{d}\tau \begin{bmatrix} \boldsymbol{a}\\ \boldsymbol{b} \end{bmatrix},$$

so

$$\Psi \Psi(\boldsymbol{a} + \imath \boldsymbol{b}) - \Psi \Psi(\boldsymbol{0}) = \int_0^1 \left(\boldsymbol{S} \boldsymbol{a} - \boldsymbol{T} \boldsymbol{b} + \imath \left(\boldsymbol{T} \boldsymbol{a} + \boldsymbol{S} \boldsymbol{b} \right) \right) \mathrm{d}\tau$$
$$= \int_0^1 \left(\boldsymbol{S} + \imath \boldsymbol{T} \right) (\boldsymbol{a} + \imath \boldsymbol{b}) \mathrm{d}\tau .$$

More generally and succintly using the Hessian definition (30.4.1):

$$\nabla \Psi(\boldsymbol{x}) = \nabla \Psi(\boldsymbol{z}) + \int_0^1 \nabla^2 \Psi(\tau \boldsymbol{x} + (1-\tau)\boldsymbol{z}) \,\mathrm{d}\tau \,(\boldsymbol{x}-\boldsymbol{z}) \,. \tag{30.4.3}$$

30.5 Convex functions of complex arguments (s,complex,convex)

Definition 29.9.7 is equally applicable to functions defined on a convex domain in \mathbb{C}^n . If $\mathcal{D} \subset \mathbb{C}^n$ is convex, then the set $\mathcal{D}_2 \triangleq \{(a, b) \in \mathbb{R}^n \times \mathbb{R}^n : a + ib \in \mathcal{D}\}$ is convex in $\mathbb{R}^n \times \mathbb{R}^n$. Clearly Ψ is convex over $\mathcal{D} \subset \mathbb{C}^n$ iff $\psi(a, b) \triangleq \Psi(a + ib)$ is convex over \mathcal{D}_2 .

Most of the properties in §29.9.3 carry over directly to convex functions on \mathbb{C}^n . The properties that require more care are those that involve derivatives or gradients. Throughout this section, we define the gradient of Ψ using (30.2.6), and we call Ψ "differentiable" on an open set $\mathcal{D} \in \mathbb{C}^n$ if $\psi(a, b) \triangleq \Psi(a + ib)$ is differentiable on \mathcal{D}_2 .

The support property (29.9.9) and its converse must be modified as follows.

Lemma 30.5.1 If Ψ is convex on $\mathcal{D} \subset \mathbb{C}^n$ and differentiable at some $z \in \mathcal{D}$, then

$$\Psi(\boldsymbol{x}) \ge \Psi(\boldsymbol{z}) + \operatorname{real}\{\langle \nabla \Psi(\boldsymbol{z}), \, \boldsymbol{x} - \boldsymbol{z} \rangle\}, \quad \forall \boldsymbol{x} \in \mathcal{D}.$$
(30.5.1)

Proof:

Using (29.9.9) and the convexity of ψ : $\Psi(\boldsymbol{x}) = \psi(\boldsymbol{x}_{\mathrm{R}}, \boldsymbol{x}_{\mathrm{I}}) \ge \psi(\boldsymbol{z}_{\mathrm{R}}, \boldsymbol{z}_{\mathrm{I}}) + \left\langle \begin{bmatrix} \nabla_{\boldsymbol{a}} \psi(\boldsymbol{z}_{\mathrm{R}}, \boldsymbol{z}_{\mathrm{I}}) \\ \nabla_{\boldsymbol{v}} \psi(\boldsymbol{z}_{\mathrm{R}}, \boldsymbol{z}_{\mathrm{I}}) \end{bmatrix}, \begin{bmatrix} \boldsymbol{x}_{\mathrm{R}} \\ \boldsymbol{x}_{\mathrm{I}} \end{bmatrix} - \begin{bmatrix} \boldsymbol{z}_{\mathrm{R}} \\ \boldsymbol{z}_{\mathrm{I}} \end{bmatrix} \right\rangle$ = $\Psi(\boldsymbol{z}) + \operatorname{real}\{\langle \nabla \Psi(\boldsymbol{z}), \, \boldsymbol{x} - \boldsymbol{z} \rangle\}.$

Lemma 30.5.2 If Ψ is differentiable on $\mathcal{D} \subset \mathbb{C}^n$, and if inequality (30.5.1) holds $\forall x, z \in \mathcal{D}$, then Ψ is convex on \mathcal{D} . Proof:

For any $x, z \in D$ and $\alpha \in [0, 1]$, let $w = \alpha x + (1 - \alpha)z$. Applying (30.5.1) twice yields

$$egin{aligned} \Psi(oldsymbol{x}) &\geq \Psi(oldsymbol{w}) + \mathrm{real}\{\langle
abla \Psi(oldsymbol{w}), oldsymbol{x} - oldsymbol{w}
angle\} \ \Psi(oldsymbol{z}) &\geq \Psi(oldsymbol{w}) + \mathrm{real}\{\langle
abla \Psi(oldsymbol{w}), oldsymbol{z} - oldsymbol{w}
angle\}. \end{aligned}$$

Thus $\alpha \Psi(\boldsymbol{x}) + (1-\alpha) \Psi(\boldsymbol{z}) \ge \Psi(\boldsymbol{w}) + \operatorname{real}\{\langle \nabla \Psi(\boldsymbol{w}), \, \alpha(\boldsymbol{x}-\boldsymbol{w}) + (1-\alpha)(\boldsymbol{z}-\boldsymbol{w}) \rangle\} = \Psi(\alpha \boldsymbol{x} + (1-\alpha) \boldsymbol{z}). \quad \Box$

onvex,line

Lemma 30.5.3 If Ψ is convex on $\mathcal{D} \subset \mathbb{C}^n$ and $\Psi(\boldsymbol{x}) = \Psi(\boldsymbol{z}) + \operatorname{real}\{\langle \nabla \Psi(\boldsymbol{z}), \boldsymbol{x} - \boldsymbol{z} \rangle\}$ where Ψ is differentiable at some $\boldsymbol{z} \in \mathcal{D}$, then for $\alpha \in [0, 1]$: $\Psi(\alpha \boldsymbol{x} + (1 - \alpha)\boldsymbol{z}) = \Psi(\boldsymbol{z}) + \alpha \operatorname{real}\{\langle \nabla \Psi(\boldsymbol{z}), \boldsymbol{x} - \boldsymbol{z} \rangle\}$. Proof:

Applying (30.5.1) at the point $\boldsymbol{w} = \alpha \boldsymbol{x} + (1 - \alpha) \boldsymbol{z}$: $\Psi(\boldsymbol{z}) + \alpha \operatorname{real}\{\langle \nabla \Psi(\boldsymbol{z}), \, \boldsymbol{x} - \boldsymbol{z} \rangle\}$ = $\Psi(\boldsymbol{z}) + \operatorname{real}\{\langle \nabla \Psi(\boldsymbol{z}), \, \boldsymbol{w} - \boldsymbol{z} \rangle\} \leq \Psi(\boldsymbol{w}) \leq \alpha \Psi(\boldsymbol{x}) + (1 - \alpha) \Psi(\boldsymbol{z})$ = $\Psi(\boldsymbol{z}) + \alpha (\Psi(\boldsymbol{x}) - \Psi(\boldsymbol{z})) = \Psi(\boldsymbol{z}) + \alpha \operatorname{real}\{\langle \nabla \Psi(\boldsymbol{z}), \, \boldsymbol{x} - \boldsymbol{z} \rangle\}.$

The next two Lemmas are similar results for strict convexity.

Lemma 30.5.4 [20, p. 9] If Ψ is strictly convex on $\mathcal{D} \subset \mathbb{C}^n$ and differentiable at some $z \in \mathcal{D}$, then

$$\Psi(\boldsymbol{x}) > \Psi(\boldsymbol{z}) + \operatorname{real}\{\langle \nabla \Psi(\boldsymbol{z}), \, \boldsymbol{x} - \boldsymbol{z} \rangle\}, \quad \forall \boldsymbol{x} \in \mathcal{D} - \{\boldsymbol{z}\}.$$
(30.5.2)

Proof:

By Lemma 30.5.1 $\Psi(\boldsymbol{x}) \geq \Psi(\boldsymbol{z}) + \operatorname{real}\{\langle \nabla \Psi(\boldsymbol{z}), \boldsymbol{x} - \boldsymbol{z} \rangle\}$. Suppose $\Psi(\boldsymbol{x}) = \Psi(\boldsymbol{z}) + \operatorname{real}\{\langle \nabla \Psi(\boldsymbol{z}), \boldsymbol{x} - \boldsymbol{z} \rangle\}$ for some $\boldsymbol{x} \neq \boldsymbol{z}$. Then Lemma 30.5.3 would hold, which would contradict the strict convexity of Ψ .

Lemma 30.5.5 If Ψ is differentiable on $\mathcal{D} \subset \mathbb{C}^n$ and (30.5.2) holds for all $x \neq z$ on \mathcal{D} , then Ψ is strictly convex. (Problem 30.5.)

Lemma 29.9.14 becomes the following generalization to \mathbb{C}^n of the monotonicity of the derivative of a convex real function.

Lemma 30.5.6 [20, p. 10] If $\Psi : \mathcal{D} \to \mathbb{R}$ is convex and differentiable on $\mathcal{D} \subset \mathbb{C}^n$, then

real{ $\langle \nabla \Psi(\boldsymbol{x}) - \nabla \Psi(\boldsymbol{z}), \, \boldsymbol{x} - \boldsymbol{z} \rangle$ } $\geq 0, \quad \forall \boldsymbol{x}, \boldsymbol{z} \in \mathcal{D}.$

Proof:

Using Lemma 30.5.1: $\Psi(\boldsymbol{z}) \geq \Psi(\boldsymbol{x}) + \operatorname{real}\{\langle \nabla \Psi(\boldsymbol{x}), \boldsymbol{z} - \boldsymbol{x} \rangle\}$ and $\Psi(\boldsymbol{x}) \geq \Psi(\boldsymbol{z}) + \operatorname{real}\{\langle \nabla \Psi(\boldsymbol{z}), \boldsymbol{x} - \boldsymbol{z} \rangle\}$, because Ψ is convex. Now add.

The following Lemmas generalize the Hessian properties (29.9.5) and (29.9.6) to the complex case. These generalizations assume that $\nabla \Psi$ is holomorphic so that $\nabla^2 \Psi$ can be defined as in (30.4.1).

Lemma 30.5.7 If Ψ is twice differentiable on \mathbb{C}^n with holomorphic gradient $\nabla \Psi$, and $\nabla^2 \Psi(\mathbf{x}) \succeq \mathbf{0}$ for all $\mathbf{x} \in \mathcal{D}$, where $\nabla^2 \Psi$ is defined in (30.4.1), then Ψ is convex on \mathcal{D} . Furthermore, if $\nabla^2 \Psi \succ \mathbf{0}$ then Ψ is strictly convex. Proof:

Using the 2nd-order Taylor series (30.4.2), when $\nabla^2 \Psi \succeq \mathbf{0}$ clearly $\Psi(\mathbf{x}) \ge \Psi(\mathbf{z}) + \operatorname{real}\{\langle \nabla \Psi(\mathbf{z}), \mathbf{x} - \mathbf{z} \rangle\}$. Now invoke Lemma 30.5.2. Similarly, when $\nabla^2 \Psi \succ \mathbf{0}$, again use the 2nd-order Taylor series (30.4.2), and invoke Lemma 30.5.5.

Lemma 30.5.8 If Ψ is twice differentiable and convex on an open convex set $\mathcal{D} \subset \mathbb{C}^n$ with holomorphic gradient $\nabla \Psi$, then $\nabla^2 \Psi(\mathbf{x}) \succeq \mathbf{0}$ for all $\mathbf{x} \in \mathcal{D}$, where $\nabla^2 \Psi$ is defined in (30.4.1),

Proof:

For $x \in D$, another form of the 2nd-order Taylor series (30.4.2) is (for ε sufficiently small):

$$\Psi(\boldsymbol{x} + \varepsilon \boldsymbol{w}) = \Psi(\boldsymbol{x}) + \operatorname{real}\{\langle \nabla \Psi(\boldsymbol{x}), \varepsilon \boldsymbol{w} \rangle\} + \frac{1}{2} (\varepsilon \boldsymbol{w})' \nabla^2 \Psi(\boldsymbol{x}) (\varepsilon \boldsymbol{w}) + \|\varepsilon \boldsymbol{w}\|^2 o(\varepsilon \boldsymbol{w}).$$

Using the inequality (30.5.1):

$$0 \leq \varepsilon^{2} \left(\frac{1}{2} \boldsymbol{w}' \, \nabla^{2} \Psi(\boldsymbol{x}) \, \boldsymbol{w} + \left\| \boldsymbol{w} \right\|^{2} o(\left\| \varepsilon \boldsymbol{w} \right\|) \right).$$

Dividing by ε^2 and letting $\varepsilon \to 0$ shows that $w' \nabla^2 \Psi(x) w \ge 0$ and because w was arbitrary, $\nabla^2 \Psi(x) \succeq 0$.

30.6 Problems (s,complex,prob)

Problem 30.1 Prove the WLS gradient expression (30.2.17).

Problem 30.2 Prove the CRB equality in (30.3.5).

Problem 30.3 In (30.2.3), we assumed that α was real. An alternative definition of the "direction of steepest ascent" that allows α to be complex might be:

$$\underset{\boldsymbol{s}:\|\boldsymbol{s}\|=1}{\operatorname{arg\,max}} \lim_{|\alpha|\to 0} \frac{\Psi(\boldsymbol{x}+\alpha\,\boldsymbol{s}) - \Psi(\boldsymbol{x})}{\alpha},$$

where α is allowed to be complex. Assuming $\psi(\mathbf{a}, \mathbf{b}) = \Psi(\mathbf{a} + i\mathbf{b})$ is differentiable, does this definition lead to a unique value for \mathbf{s} ? If so, compare to (30.2.3). (Need typed.)

Problem 30.4 Consider the model $\bar{y}(x)$ defined by $\bar{y}_i(x) = f_i e^{x_i}$ where $f_i, x_i \in \mathbb{C}$ and f_i are known constants. (This model arises in certain MRI applications where one is estimating relaxation parameters and phase parameters.) Determine whether $\bar{y}(x)$ is holomorphic, and if so, find an expression for $\nabla \Psi(x)$.

Problem 30.5 Prove Lemma 30.5.5 that relates (30.5.2) to strict convexity.

30.7 Bibliography

olhede:06:opd

kim:20:eis

- D. H. Brandwood. "A complex gradient operator and its application in adaptive array theory." In: *IEE Proceedings H: Microwaves, Optics and Antennas* 130.1 (Feb. 1983), 11–6. DOI: 10.1049/ip-f-1:19830003 (cit. on p. 30.1).
- [2] A. van den Bos. "Complex gradient and Hessian." In: *IEE Proceedings Vision, Image and Signal Processing* 141.6 (Dec. 1994), 380–3. DOI: 10.1049/ip-vis:19941555 (cit. on p. 30.1).
- [3] S. C. Olhede. "On probability density functions for complex variables." In: *IEEE Trans. Info. Theory* 52.3 (Mar. 2006), 1212–7. DOI: 10.1109/TIT.2005.864451 (cit. on p. 30.1).
- [4] A. Hjorungnes and D. Gesbert. "Complex-valued matrix differentiation: techniques and key results." In: *IEEE Trans. Sig. Proc.* 55.6 (June 2007), 2740–6. DOI: 10.1109/TSP.2007.893762 (cit. on p. 30.1).
- [5] K. Kreutz-Delgado. *The complex gradient operator and the CR-calculus*. 2009. URL: http://arxiv.org/abs/0906.4835 (cit. on p. 30.1).
- [6] T. H. Kim and J. P. Haldar. *Efficient iterative solutions to complex-valued nonlinear least-squares problems* with mixed linear and antilinear operators. 2020. URL: http://arxiv.org/abs/2007.09281 (cit. on p. 30.1).
- [7] R. V. Churchill. *Complex variables and applications*. 2nd edition. New York: McGraw-Hill, 1960 (cit. on p. 30.1).
- [8] D. G. Luenberger. *Optimization by vector space methods*. New York: Wiley, 1969. URL: http://books.google.com/books?id=lZU0CAH4RccC (cit. on p. 30.2).
- [9] T. Adali, P. J. Schreier, and L. L. Scharf. "Complex-valued signal processing: the proper way to deal with impropriety." In: *IEEE Trans. Sig. Proc.* 59.11 (Nov. 2011), 5101–25. DOI: 10.1109/TSP.2011.2162954 (cit. on p. 30.2).

(Need typed.)

sorber:12:uod

- [10] E. Candes, X. Li, and M. Soltanolkotabi. *Phase retrieval via Wirtinger flow: Theory and algorithms*. 2014. URL: http://arxiv.org/abs/1407.1065 (cit. on p. 30.2).
- [11] L. Sorber, M. V. Barel, and L. D. Lathauwer. "Unconstrained optimization of real functions in complex variables." In: *SIAM J. Optim.* 22.3 (2012), 879–98. DOI: 10.1137/110832124 (cit. on p. 30.3).
- [12] J. Liu et al. "First- and second-order methods for online convolutional dictionary learning." In: SIAM J. Imaging Sci. 11.2 (Jan. 2018), 1589–628. DOI: 10.1137/17m1145689 (cit. on p. 30.3).
- G. Sebastiani and F. Godtliebsen. "On the use of Gibbs priors for Bayesian image restoration." In: *Signal Processing* 56.1 (Jan. 1997), 111–8. DOI: 10.1016/S0165–1684 (97) 00002–9 (cit. on p. 30.3).
- [14] V. Olafsson, J. A. Fessler, and D. C. Noll. "Spatial resolution analysis of iterative image reconstruction with separate regularization of real and imaginary parts." In: *Proc. IEEE Intl. Symp. Biomed. Imag.* 2006, 5–8. DOI: 10.1109/ISBI.2006.1624838 (cit. on p. 30.3).
- [15] W. S. Hoge et al. "Fast regularized reconstruction of non-uniformly subsampled partial-Fourier parallel MRI data." In: *Proc. IEEE Intl. Symp. Biomed. Imag.* 2007, 1012–5. DOI: 10.1109/ISBI.2007.357026 (cit. on p. 30.3).
- [16] L. Chen, M. C. Schabel, and E. V. R. DiBella. "Reconstruction of dynamic contrast enhanced magnetic resonance imaging of the breast with temporal constraints." In: *Mag. Res. Im.* 28.5 (June 2010), 637–45. DOI: 10.1016/j.mri.2010.03.001 (cit. on p. 30.3).
 - [17] P. Joly and Gerard Meurant. "Complex conjugate gradient methods." In: *Numer. Algorithms* 4.3 (1993), 379–406. DOI: 10.1007/BF02145754 (cit. on p. 30.6).
 - [18] J. D. Willig-Onwuachi et al. "Phase-constrained parallel MR image reconstruction." In: J. Mag. Res. 176.2 (Oct. 2005), 187–98. DOI: 10.1016/j.jmr.2005.06.004 (cit. on pp. 30.6, 30.7).
 - [19] A. van den Bos. "A Cramer-Rao lower bound for complex parameters." In: *IEEE Trans. Sig. Proc.* 42.10 (Oct. 1994), p. 2859. DOI: 10.1109/78.324755 (cit. on p. 30.7).
 - [20] B. T. Polyak. Introduction to optimization. New York: Optimization Software Inc, 1987. URL: http://lab7.ipu.ru/eng/sel-papers.html (cit. on p. 30.10).