Minimization of sums of squares of holomorphic functions

Roland Hildebrand

Laboratoire Jean Kuntzmann / CNRS

Optimization without borders: Huawei Day

Sochi, July 15, 2021

Problem statement

consider the optimization problem

$$\min_{z \in X} f(z)$$

where $f(z) = \sum_{k=1}^{K} |g_k(z)|^2$, g_k holomorphic in the domain $X \subset \mathbb{C}^n$

regard $\mathbb{C}^n$ as a real space $\mathbb{R}^{2n}$

then we may tackle the problem as minimization of $f(z) = \sum_{k=1}^{K} \left((Re\, g_k(z))^2 + (Im\, g_k(z))^2\right)$ with gradient descent, Newton etc.

Can we use the complex structure to facilitate calculations and / or accelerate convergence?

Motivation

Huawei observed that the sequence produced by the "mixed Newton" iteration

$$z_{j+1} = z_j - \left(\frac{\partial^2 f}{\partial \bar{z} \partial z} \right)^{-1} \frac{\partial f}{\partial \bar{z}}$$

converges faster than the real gradient descent method and better avoids saddle points

several examples have been tested

- $K = n = 1$, and $g_1(z) = g(z)$ is a polynomial of a scalar variable z
- $z = (u, v)$, $g_k(z) = (a_k^T u) \cdot (b_k^T v)$, where a_k, b_k are complex vectors of appropriate size

Wirtinger calculus

calculations of gradients and Hessians in the real variables $Re\, z$, $Im\, z$ are comparatively laborous

but a *holomorphic* function g is completely determined by its *complex* derivatives $\frac{\partial^k g}{\partial z^k}$, which are also easier to compute

Wirtinger calculus describes how to differentiate functions f depending on holomorphic functions g_k and their conjugates using only complex derivatives

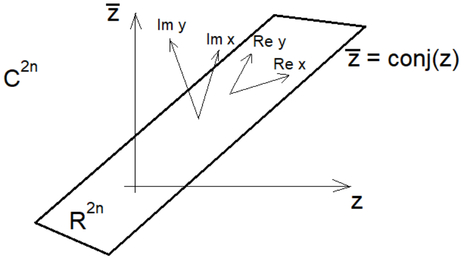
due to the conjugates we need an additional complex variable $\bar{z}$

the variables $z, \bar{z}$ depend complex-linearly on the complexified real variables $x = Re\, z$, $y = Im\, z$:

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} z \\ \bar{z} \end{pmatrix} = \begin{pmatrix} I & iI \\ I & -iI \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

Wirtinger calculus

embed $\mathbb{C}^n \sim \mathbb{R}^{2n}$ into $\mathbb{C}^{2n}$ by complexification of x, y
use variables $z, \bar{z}$ on $\mathbb{C}^{2n}$



how to extend f from $\mathbb{R}^{2n}$ to $\mathbb{C}^{2n}$?

- express f as a function of holomorphic functions g_k and their conjugates $\bar{g}_k$
- g_k are assumed independent of $\bar{z}$
- $\bar{g}_k$ are assumed independent of z

Wirtinger calculus

the partial derivatives

$$\frac{\partial(z,\bar{z})}{\partial(x,y)} = \begin{pmatrix} I & iI \\ I & -iI \end{pmatrix},$$

$$\frac{\partial(x,y)}{\partial(z,\bar{z})} = \frac{1}{2}\begin{pmatrix} I & I \\ -iI & iI \end{pmatrix}$$

are constant

we get explicit expressions for the differentiation operators

$$\frac{\partial}{\partial z} = \frac{1}{2}\left(\frac{\partial}{\partial x} - i\frac{\partial}{\partial y}\right), \quad \frac{\partial}{\partial \bar{z}} = \frac{1}{2}\left(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\right)$$

Wirtinger calculus consists of a set of rules to compute derivatives with respect to $z, \bar{z}$ without resorting to x, y

Wirtinger calculus

some rules

- g holomorphic: $\frac{\partial g}{\partial z} = g'$, $\frac{\partial g}{\partial \bar{z}} = 0$
- g holomorphic: $\frac{\partial \bar{g}}{\partial z} = 0$, $\frac{\partial \bar{g}}{\partial \bar{z}} = \bar{g}'$
- g holomorphic or anti-holomorphic: mixed (containing entries of z and $\bar{z}$) derivatives vanish
- $\frac{\partial}{\partial z}(af + bg) = a\frac{\partial f}{\partial z} + b\frac{\partial g}{\partial z}$, $a, b \in \mathbb{C}$
- $\frac{\partial}{\partial z}(f \cdot g) = f\frac{\partial g}{\partial z} + g\frac{\partial f}{\partial z}$
- $\frac{\partial}{\partial z}(f \circ g) = \left(\frac{\partial f}{\partial w} \circ g\right) \cdot \frac{\partial g}{\partial z} + \left(\frac{\partial f}{\partial \bar{w}} \circ g\right) \cdot \frac{\partial \bar{g}}{\partial z}$
- $\overline{\frac{\partial g}{\partial z}} = \frac{\partial \bar{g}}{\partial \bar{z}}$

Sum of squares of holomorphic functions

let $f(z) = \sum_{k=1}^{K} |g_k(z)|^2$, g_k holomorphic

express $f = \sum_{k=1}^{K} g_k \cdot \bar{g}_k$, here g_k depends on z only, and $\bar{g}_k$ depends on $\bar{z}$ only

the derivatives are given by

$$\frac{\partial f}{\partial z} = \sum_{k=1}^{K} \bar{g}_k \cdot \frac{\partial g_k}{\partial z} = \sum_{k=1}^{K} \bar{g}_k \cdot g_k', \quad \frac{\partial f}{\partial \bar{z}} = \sum_{k=1}^{K} g_k \cdot \frac{\partial \bar{g}_k}{\partial \bar{z}} = \sum_{k=1}^{K} g_k \cdot \overline{g_k'}$$

$$\frac{\partial^2 f}{\partial z^2} = \sum_{k=1}^{K} \bar{g}_k \cdot g_k'', \qquad \frac{\partial^2 f}{\partial z \partial \bar{z}} = \sum_{k=1}^{K} g_k' \cdot (\overline{g_k'})^T$$

$$\frac{\partial^2 f}{\partial \bar{z} \partial z} = \sum_{k=1}^{K} \overline{g_k'} \cdot (g_k')^T, \qquad \frac{\partial^2 f}{\partial \bar{z}^2} = \sum_{k=1}^{K} \sum_{k=1}^{K} g_k \cdot \overline{g_k''}$$

mixed derivatives positive semi-definite hermitian of rank K in general

Taylor approximation

- gradient descent is moving along the descent direction of the linear approximation
- Newton method is minimizing the quadratic approximation
- in the real case the approximations are obtained from the Taylor expansion

Taylor expansion in complex variables

$$
\begin{aligned}
f(z+h) =& f(z) + \left\langle \frac{\partial f}{\partial z}, h \right\rangle_{\mathbb{R}} + \left\langle \frac{\partial f}{\partial \bar{z}}, \bar{h} \right\rangle_{\mathbb{R}} + \\
&+ \frac{1}{2} \begin{pmatrix} h \\ \bar{h} \end{pmatrix}^T \begin{pmatrix} \frac{\partial^2 f}{\partial z^2} & \frac{\partial^2 f}{\partial z \partial \bar{z}} \\ \frac{\partial^2 f}{\partial \bar{z} \partial z} & \frac{\partial^2 f}{\partial \bar{z}^2} \end{pmatrix} \begin{pmatrix} h \\ \bar{h} \end{pmatrix} + o(|h|^2)
\end{aligned}
$$

here $\langle u, v \rangle_{\mathbb{R}} = \sum_i u_i v_i = \langle u, \bar{v} \rangle$

Differences with real case

linear term

steepest descent: $\left\langle \frac{\partial f}{\partial z}, \bar{h} \right\rangle + \left\langle \frac{\partial f}{\partial \bar{z}}, h \right\rangle$ minimal $\Leftrightarrow$ $h, \frac{\partial f}{\partial \bar{z}}$ anti-parallel

quadratic term

real case: convexity / concavity determined by signature of Hessian
$f'' = \left(\frac{\partial^2 f}{\partial x_i \partial x_j} \right)_{ij}$
complex case: signature makes sense only for *hermitian* matrices
symmetric complex matrices may have complex eigenvalues

real case: quadratic form evaluates on vectors by $x \mapsto x^T A x$

complex case: hermitian form evaluates on vectors by $z \mapsto z^* H z$

we have to exchange $h, \bar{h}$ in one of the tangent vectors at the Hessian

Transformed Hessian

applying the transformations we get

$$f(z+h) = f(z) + \left\langle \frac{\partial f}{\partial z}, \bar{h} \right\rangle + \left\langle \frac{\partial f}{\partial \bar{z}}, h \right\rangle + \\ + \frac{1}{2} \begin{pmatrix} h \\ \bar{h} \end{pmatrix}^* \begin{pmatrix} \frac{\partial^2 f}{\partial \bar{z} \partial z} & \frac{\partial^2 f}{\partial \bar{z}^2} \\ \frac{\partial^2 f}{\partial z^2} & \frac{\partial^2 f}{\partial z \partial \bar{z}} \end{pmatrix} \begin{pmatrix} h \\ \bar{h} \end{pmatrix} + o(|h|^2)$$

the matrix in the second-order term is now hermitian

for $f = \sum_{k=1}^{K} |g_k|^2$ with g_k holomorphic the diagonal blocks are *always* positive semi-definite
if $K \geq n$, then in general the diagonal blocks are positive definite but this holds only on $\mathbb{R}^{2n} \subset \mathbb{C}^{2n}$

Convexity and signature

the quadratic part of f on the subspace $\mathbb{R}^{2n}$ defined by $x, y \in \mathbb{R}^n$ is given by the real symmetric matrix

$$\begin{pmatrix} I & I \\ -iI & iI \end{pmatrix} \begin{pmatrix} \frac{\partial^2 f}{\partial \bar{z} \partial z} & \frac{\partial^2 f}{\partial \bar{z}^2} \\ \frac{\partial^2 f}{\partial z^2} & \frac{\partial^2 f}{\partial z \partial \bar{z}} \end{pmatrix} \begin{pmatrix} I & iI \\ I & -iI \end{pmatrix} =$$

$$\begin{pmatrix} \frac{\partial^2 f}{\partial \bar{z}\partial z} + \frac{\partial^2 f}{\partial z\partial \bar{z}} + \frac{\partial^2 f}{\partial \bar{z}^2} + \frac{\partial^2 f}{\partial z^2} & i\left(\frac{\partial^2 f}{\partial \bar{z}\partial z} - \frac{\partial^2 f}{\partial z\partial \bar{z}} + \frac{\partial^2 f}{\partial z^2} - \frac{\partial^2 f}{\partial \bar{z}^2}\right) \\ i\left(-\frac{\partial^2 f}{\partial \bar{z}\partial z} + \frac{\partial^2 f}{\partial z\partial \bar{z}} + \frac{\partial^2 f}{\partial z^2} - \frac{\partial^2 f}{\partial \bar{z}^2}\right) & \frac{\partial^2 f}{\partial \bar{z}\partial z} + \frac{\partial^2 f}{\partial z\partial \bar{z}} - \frac{\partial^2 f}{\partial \bar{z}^2} - \frac{\partial^2 f}{\partial z^2} \end{pmatrix}$$

in coordinates $\begin{pmatrix} Re\, h \\ Im\, h \end{pmatrix} = \begin{pmatrix} I & iI \\ I & -iI \end{pmatrix}^{-1} \begin{pmatrix} h \\ \bar{h} \end{pmatrix}$

signature of these matrices determines convexity / concavity on $\mathbb{R}^{2n}$

at least n positive eigenvalues: no maxima, "maximal" saddle-point has neutral signature

Real quadratic approximation

this Taylor approximation is valid only on $\mathbb{R}^{2n} \sim \mathbb{C}^n$, where $\bar{h}$ is the conjugate of h

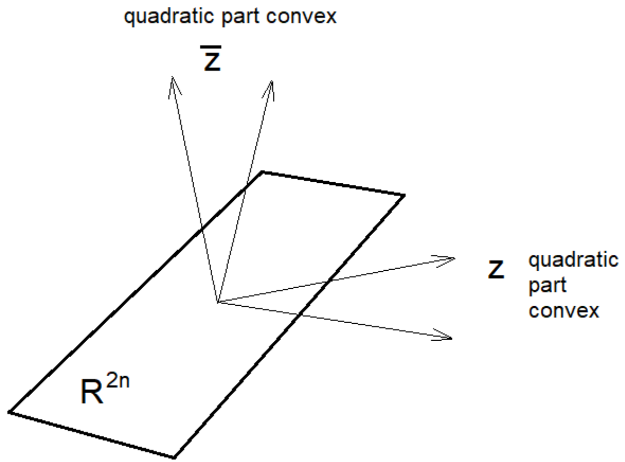
set $\mathbf{h} = (h^T, \bar{h}^T)^T$, $\mathbf{z} = (z^T, \bar{z}^T)^T$
here $z, \bar{z}$ are conjugate, but $h, \bar{h}$ are treated as independent variables
then $\mathbf{z} + \mathbf{h}$ parameterizes $\mathbb{C}^{2n}$

define the real quadratic approximation

$$q_z(\mathbf{z} + \mathbf{h}) = f(z) + Re \left\langle \begin{pmatrix} \frac{\partial f}{\partial \bar{z}} \\ \frac{\partial f}{\partial z} \end{pmatrix}, \mathbf{h} \right\rangle + \frac{1}{2} \mathbf{h}^* \begin{pmatrix} \frac{\partial^2 f}{\partial \bar{z} \partial z} & \frac{\partial^2 f}{\partial \bar{z}^2} \\ \frac{\partial^2 f}{\partial z^2} & \frac{\partial^2 f}{\partial z \partial \bar{z}} \end{pmatrix} \mathbf{h}$$

differs from complex second order Taylor polynomial outside of $\mathbb{R}^{2n}$
convex on $\mathbb{R}^{2n}$ if and only if convex on $\mathbb{C}^{2n}$

Definiteness of diagonal blocks



q_z convex on the real $2n$-planes spanned by the variables $z, \bar{z}$ in $\mathbb{C}^{2n}$

First interpretation of "mixed Newton"

consider the real quadratic function

$$q(\mathbf{z}+\mathbf{h}) = f(z) + Re\langle v, \mathbf{h}\rangle + \frac{1}{2}\mathbf{h}^* A\mathbf{h}$$

with $v \in \mathbb{C}^n$, A complex hermitian invertible of size $2n$

the unique stationary point of q is given by $\mathbf{z} - A^{-1}v$

the "mixed Newton" method yields the stationary point of the *diagonalised-convexified* quadratic approximation

$$\tilde{q}_z(\mathbf{z}+\mathbf{h}) = f(z) + Re\left\langle \begin{pmatrix} \frac{\partial f}{\partial \bar{z}} \\ \frac{\partial f}{\partial z} \end{pmatrix}, \mathbf{h} \right\rangle + \frac{1}{2}\mathbf{h}^* \begin{pmatrix} \frac{\partial^2 f}{\partial \bar{z}\partial z} & \mathbf{0} \\ \mathbf{0} & \frac{\partial^2 f}{\partial z \partial \bar{z}} \end{pmatrix} \mathbf{h}$$

$$\mathbf{z} \mapsto \mathbf{z} - \begin{pmatrix} \left(\frac{\partial^2 f}{\partial \bar{z}\partial z}\right)^{-1} \frac{\partial f}{\partial \bar{z}} \\ \left(\frac{\partial^2 f}{\partial z\partial \bar{z}}\right)^{-1} \frac{\partial f}{\partial z} \end{pmatrix}, \quad z \mapsto z - \left(\frac{\partial^2 f}{\partial \bar{z}\partial z}\right)^{-1} \frac{\partial f}{\partial \bar{z}}$$

the iteration stays in $\mathbb{R}^{2n}$

First interpretation of "mixed Newton"

- extend function f from $\mathbb{R}^{2n}$ to $\mathbb{C}^{2n}$, complex-valued
- build quadratic Taylor approximation, complex-valued
- extend quadratic approximation from $\mathbb{R}^{2n}$ to $\mathbb{C}^{2n}$, real-valued
- drop off-diagonal blocks, diagonalised-convexified quadratic approximation real-valued and convex
- go to stationary point of diagonalised-convexified quadratic approximation

gradient descent with block-diagonal preconditioning by inverse of modified Hessian
block-diagonal in complex space — no partition of real coordinates in subsets

Special case: scalar z

let $n = \dim^{\mathbb{C}} z = 1$
then the entries of the "Hermitian Hessian" are scalars

in the "mixed Newton" iteration

$$z_{j+1} = z_j - \left(\frac{\partial^2 f}{\partial \bar{z} \partial z} \right)^{-1} \frac{\partial f}{\partial \bar{z}}$$

the inverse of $\frac{\partial^2 f}{\partial \bar{z} \partial z}$ is just a positive step length

the 2-real-dimensional *direction* of the step is still given by the complex number $\frac{\partial f}{\partial \bar{z}}$
this is just the steepest descent direction

"mixed Newton" is gradient descent with an adjusted step size

Very special case: one holomorphic function

suppose $n = K = 1$

we look just for the zeros of $g(z)$

$$\frac{\partial^2 f}{\partial \bar{z} \partial z} = |g'|^2, \ \frac{\partial f}{\partial \bar{z}} = g\bar{g}' : \quad z_{j+1} = z_j - \frac{g(z_j)}{g'(z_j)}$$

this is the ordinary Newton method when interpreting g as a vector field on $\mathbb{C} \sim \mathbb{R}^2$

$$z_{j+1} - z^* \approx \frac{g''(z^*) \cdot (z_j - z^*)^2}{2g'(z_j)}$$

- quadratic convergence if $g'(z^*) \neq 0$
- linear convergence if zero of g is multiple: $z_{j+1} - z^* \approx \frac{m-1}{m}(z_j - z^*)$

Special case: linear g_k

then $\frac{\partial^2 f}{\partial z^2} = \frac{\partial^2 f}{\partial \bar{z}^2} = \mathbf{0}$

- modified quadratic approximation coincides with real quadratic approximation q_z
- approximation q_z convex on $\mathbb{C}^{2n}$, in particular, on $\mathbb{R}^{2n}$
- "mixed Newton" step coincides with ordinary real Newton step in $\mathbb{R}^{2n}$

- quadratic Taylor approximation coincides with original cost function f
- global minimum of f achieved at first step

Second interpretation of "mixed Newton"

- approximate holomorphic functions g_k by complex-affine functions $l_k(z) = a_k^* z + b_k$
- go to global minimum of $\sum_{k=1}^{K} |l_k(z)|^2$

explicit formula

$$z_{j+1} = z_j - \left(\sum_{k=1}^{K} a_k a_k^* \right)^{-1} \sum_{k=1}^{K} l_k(z) a_k$$

Convergence analysis

let $z^* \in \mathbb{C}^n$ be a local minimum of f, $z = z^* + \delta$

denote $\gamma_k = g_k(z^*)$, $\nabla_k = g'_k(z^*)$, $M_k = g''_k(z^*)$, then

$$g'_k(z) = \nabla_k + M_k\delta + O(|\delta|^2), \; g_k(z) = \gamma_k + \nabla_k^T\delta + O(|\delta|^2)$$

$$\frac{\partial^2 f}{\partial \bar{z}\partial z} = \sum_{k=1}^{K} \bar{\nabla}_k \nabla_k^T + O(|\delta|)$$

$$\frac{\partial f}{\partial \bar{z}} = \sum_{k=1}^{K} (\gamma_k + \nabla_k^T\delta + O(|\delta|^2))(\bar{\nabla}_k + \bar{M}_k\bar{\delta} + O(|\delta|^2))$$

$$= \sum_{k=1}^{K} \left(\gamma_k \bar{\nabla}_k + \gamma_k \bar{M}_k \bar{\delta} + \bar{\nabla}_k \nabla_k^T \delta \right) + O(|\delta|^2)$$

$$= \sum_{k=1}^{K} \left(\gamma_k \bar{M}_k \bar{\delta} + \bar{\nabla}_k \nabla_k^T \delta \right) + O(|\delta|^2)$$

$\sum_{k=1}^{K} \gamma_k \bar{\nabla}_k = 0$ by first order optimality condition

Convergence analysis

the "mixed Newton" step operates by

$$\begin{aligned}\delta \mapsto \delta &- \left(\sum_{k=1}^{K} \bar{\nabla}_k \nabla_k^T\right)^{-1} \left(\sum_{k=1}^{K} \left(\gamma_k \bar{M}_k \bar{\delta} + \bar{\nabla}_k \nabla_k^T \delta\right)\right) + O(|\delta|^2) \\ &= - \left(\sum_{k=1}^{K} \bar{\nabla}_k \nabla_k^T\right)^{-1} \left(\sum_{k=1}^{K} \gamma_k \bar{M}_k\right) \bar{\delta} + O(|\delta|^2) \\ &= -H_{11}^{-1} H_{12} \bar{\delta} + O(|\delta|^2)\end{aligned}$$

where $H = \frac{1}{2} \begin{pmatrix} H_{11} & H_{12} \\ \bar{H}_{12} & \bar{H}_{11} \end{pmatrix}$ determines the quadratic term of q_{z^*}

two iterations give
$\delta \mapsto -H_{11}^{-1} H_{12} \bar{\delta} + O(|\delta|^2) \mapsto H_{11}^{-1} H_{12} \bar{H}_{11}^{-1} \bar{H}_{12} \delta + O(|\delta|^2)$

linearly convergent if and only if $\mathrm{spec}(H_{11}^{-1} H_{12} \bar{H}_{11}^{-1} \bar{H}_{12})$ in the open unit disc D^o

Convergence analysis

Lemma

Let $H_{11} \succ 0$. *Then* $\mathrm{spec}(H_{11}^{-1}H_{12}\bar{H}_{11}^{-1}\bar{H}_{12}) \subset D^o$ *if and only if* $H \succ 0$.

proof:

- $\mathrm{spec}(H_{11}^{-1}H_{12}\bar{H}_{11}^{-1}\bar{H}_{12}) = \mathrm{spec}(H_{11}^{-1/2}H_{12}\bar{H}_{11}^{-1}\bar{H}_{12}H_{11}^{-1/2}) \subset \mathbb{R}_{++}$
- $\lambda_{\max}(H_{11}^{-1/2}H_{12}\bar{H}_{11}^{-1}\bar{H}_{12}H_{11}^{-1/2}) = \sigma_{\max}(H_{11}^{-1/2}H_{12}\bar{H}_{11}^{-1/2})^2$
- $\sigma_{\max}(H_{11}^{-1/2}H_{12}\bar{H}_{11}^{-1/2}) < 1$ if and only if $\begin{pmatrix} I & H_{11}^{-1/2}H_{12}\bar{H}_{11}^{-1/2} \\ \bar{H}_{11}^{-1/2}\bar{H}_{12}H_{11}^{-1/2} & I \end{pmatrix} \succ 0$
- if and only if $H \succ 0$

each negative eigenvalue of H yields a repulsive direction for the dynamics

Behaviour at critical points

if $K \geq n$ then generically the "mixed Newton" iteration is

- linearly convergent in the neighbourhood of local minima
- follows the trajectories of a hyperbolic system at "non-maximal" saddle points
- repulsive at "maximal" saddle-points (neutral signature of "Hermitian Hessian")

Huawei problem with bi-linear g_k:

- all g_k' share a common orthogonal direction corresponding to the symmetry $(u, v) \mapsto (\alpha u, \alpha^{-1} v)$, $\alpha \in \mathbb{C} \setminus \{0\}$
- diagonal blocks $\frac{\partial^2 f}{\partial \bar{z} \partial z}$ have a kernel vector
- matrix of dynamic system has an eigenvalue 1

Thank you!