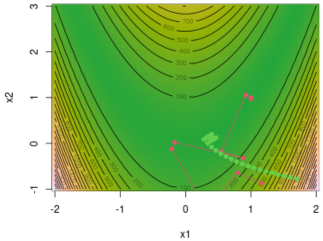
Optimization in Machine Learning

Second order methods
Quasi-Newton



Learning goals

- Newton-Raphson vs. Quasi-Newton
- SR1
- BFGS

QUASI-NEWTON: IDEA

Start point of **QN method** is (as with NR) a Taylor approximation of the gradient, except that H is replaced by a **pd** matrix $\boldsymbol{A}^{[t]}$:

$$
\begin{aligned}
\nabla f(\mathbf{x}) &\approx \nabla f(\mathbf{x}^{[t]}) + \nabla^2 f(\mathbf{x}^{[t]})(\mathbf{x} - \mathbf{x}^{[t]}) = \mathbf{0} && \text{NR} \\
\nabla f(\mathbf{x}) &\approx \nabla f(\mathbf{x}^{[t]}) + \boldsymbol{A}^{[t]} \quad (\mathbf{x} - \mathbf{x}^{[t]}) = \mathbf{0} && \text{QN}
\end{aligned}
$$

The update direction:

$$
\begin{aligned}
\boldsymbol{d}^{[t]} &= -\nabla^2 f(\mathbf{x}^{[t]})^{-1} \nabla f(\mathbf{x}^{[t]}) && \text{NR} \\
\boldsymbol{d}^{[t]} &= -(\boldsymbol{A}^{[t]})^{-1} \quad \nabla f(\mathbf{x}^{[t]}) && \text{QN}
\end{aligned}
$$

SYMMETRIC RANK 1 UPDATE (SR1)

Simplest approach: symmetric rank 1 updates (**SR1**) of form

$$\mathbf{A}^{[t+1]} \leftarrow \mathbf{A}^{[t]} + \mathbf{B}^{[t]} = \mathbf{A}^{[t]} + \beta \mathbf{u}^{[t]} (\mathbf{u}^{[t]})^\top$$

with appropriate vector $\mathbf{u}^{[t]} \in \mathbb{R}^n$, $\beta \in \mathbb{R}$.

BFGS ALGORITHM

Instead of Rank 1 updates, the **BFGS** procedure (published simultaneously in 1970 by Broyden, Fletcher, Goldfarb and Shanno) uses rank 2 modifications of the form

$$\mathbf{A}^{[t]} + \beta_1 \mathbf{u}^{[t]} (\mathbf{u}^{[t]})^\top + \beta_2 \mathbf{v}^{[t]} (\mathbf{v}^{[t]})^\top$$

with $\mathbf{s}^{[t]} := \mathbf{x}^{[t+1]} - \mathbf{x}^{[t]}$

- $\mathbf{u}^{[t]} = \nabla f(\mathbf{x}^{[t+1]}) - \nabla f(\mathbf{x}^{[t]})$
- $\mathbf{v}^{[t]} = \mathbf{A}^{[t]} \mathbf{s}^{[t]}$
- $\beta_1 = \frac{1}{(\mathbf{u}^{[t]})^\top (\mathbf{s}^{[t]})}$
- $\beta_2 = -\frac{1}{(\mathbf{s}^{[t]})^\top \mathbf{A}^{[t]} \mathbf{s}^{[t]}}$

The resulting matrices $\mathbf{A}^{[t]}$ are positive definite and the corresponding quasi-newton update directions $\mathbf{d}^{[t]}$ are actual descent directions.