# Low-Rank Matrix Updates

- Problem Formulation
- Preliminaries
- Broyden Updates
- BFGS Updates

# Contents

# Problem Formulation

We have learned that Newton type methods for (unconstrained) optimization proceed by implementing iterates of the form

$$x^+ = x - M(x)^{-1} F'(x)^T \; .$$

Here, $M(x) \in \mathbb{R}^{n \times n}$ is an approximation of $F''(x)$.

Problem: Can we construct "cheap" approximation of $F''(x)$ such that

- we don't have to evaluate second order derivatives and
- we can cheapy compute $M(x)^{-1}$ even if $n$ is large?

# Problem Formulation

We have learned that Newton type methods for (unconstrained) optimization proceed by implementing iterates of the form

$$x^+ = x - M(x)^{-1} F'(x)^T \ .$$

Here, $M(x) \in \mathbb{R}^{n \times n}$ is an approximation of $F''(x)$.

Problem: Can we construct "cheap" approximation of $F''(x)$ such that

- we don't have to evaluate second order derivatives and
- we can cheapy compute $M(x)^{-1}$ even if $n$ is large?

# Contents

# Low-rank matrices

Storing "big" matrices of the form $A \in \mathbb{R}^{n \times n}$ can be a problem if $n$ is large. One exception are matrices that can be represented in the form

$$A = UV^T$$

with $U, V \in \mathbb{R}^{n \times m}$ where $m \ll n$. Matrices of this form are not invertible and are called low-rank matrices.

An important special case is obtained for $m = 1$, where $U$ and $V$ are vectors, which yields rank-1 matrices.

# Low-rank matrices

Storing "big" matrices of the form $A \in \mathbb{R}^{n\times n}$ can be a problem if $n$ is large. One exception are matrices that can be represented in the form

$$A = UV^T$$

with $U, V \in \mathbb{R}^{n\times m}$ where $m \ll n$. Matrices of this form are not invertible and are called low-rank matrices.

An important special case is obtained for $m = 1$, where $U$ and $V$ are vectors, which yields rank-1 matrices.

# Woodbury's matrix inversion formula

One way to represent invertible matrices is by considering matrices of the form

$$A = B + UV^T$$

where $B \in \mathbb{R}^{n\times n}$ is an "easy-to-store" matrix that is invertible and $U, V \in \mathbb{R}^{n\times m}$.

If the matrix $B$ is easy to invert (or we know $B^{-1}$ already), the inverse of the matrix $A$ can be found from

$$\left(B + UV^T\right)^{-1} = B^{-1} - B^{-1}U\left(I + V^TB^{-1}U\right)^{-1}V^TB^{-1}$$

# Woodbury's matrix inversion formula

One way to represent invertible matrices is by considering matrices of the form

$$A = B + UV^T$$

where $B \in \mathbb{R}^{n\times n}$ is an "easy-to-store" matrix that is invertible and $U, V \in \mathbb{R}^{n\times m}$.

If the matrix $B$ is easy to invert (or we know $B^{-1}$ already), the inverse of the matrix $A$ can be found from

$$\left(B + UV^T\right)^{-1} = B^{-1} - B^{-1}U\left(I + V^TB^{-1}U\right)^{-1}V^TB^{-1}$$

# Woodbury's matrix inversion formula

The inversion formula

$$\left(B + UV^T\right)^{-1} = B^{-1} - B^{-1}U\left(I + V^TB^{-1}U\right)^{-1}V^TB^{-1}$$

is known under the name Woodbury formula (or Sherman-Morrison-Woodbury formula).

For the special case $m = 1$ the matrix $\left(I + V^TB^{-1}U\right)$ is scalar and can be trivially inverted. In general, we only need to invert an $(m \times m)$-matrix instead of a $(n \times n)$-matrix.

# Woodbury's matrix inversion formula

The inversion formula

$$\left(B + UV^T\right)^{-1} = B^{-1} - B^{-1}U\left(I + V^TB^{-1}U\right)^{-1}V^TB^{-1}$$

is known under the name Woodbury formula (or Sherman-Morrison-Woodbury formula).

For the special case $m = 1$ the matrix $\left(I + V^TB^{-1}U\right)$ is scalar and can be trivially inverted. In general, we only need to invert an $(m \times m)$-matrix instead of a $(n \times n)$-matrix.

# Proof of Woodbury's matrix inversion formula

A proof Woodbury's matrix inversion formula can be obtained by direct verification:

$$
\begin{aligned}
&\left(B + UV^T\right)\left(B^{-1} - B^{-1}U\left(I + V^TB^{-1}U\right)^{-1}V^TB^{-1}\right) \\
&= I + UV^TB^{-1} - \left(U + UV^TB^{-1}U\right)\left(I + V^TB^{-1}U\right)^{-1}V^TB^{-1} \\
&= I + UV^TB^{-1} - U\left(I + V^TB^{-1}U\right)\left(I + V^TB^{-1}U\right)^{-1}V^TB^{-1} \\
&= I + UV^TB^{-1} - UV^TB^{-1} = I \ .
\end{aligned}
$$

# Some other useful results from matrix analysis

The derivative of a function $f : \mathbb{R}^{n\times n} \to D$ in the direction $\Delta \in \mathbb{R}^{n\times n}$ can be defined as

$$\frac{\partial f(X)}{\partial X} \circ \Delta = \lim_{h\to 0} \frac{f(X + h\Delta) - f(X)}{h}$$

Important examples:

- $\frac{\partial \mathrm{Tr}(AX)}{\partial X} \circ \Delta = \mathrm{Tr}(A\Delta)$.
- $\frac{\partial X^{-1}}{\partial X} \circ \Delta = -X^{-1}\Delta X^{-1}$.
- $\frac{\partial \mathrm{Tr}(AXBX^TC)}{\partial X} \circ \Delta = \mathrm{Tr}(A\Delta BX^TC + AXB\Delta^TC) = \mathrm{Tr}(\left[BX^TCA + B^TX^TC^TA^T\right]\Delta)$.

# Some other useful results from matrix analysis

The derivative of a function $f : \mathbb{R}^{n\times n} \to D$ in the direction $\Delta \in \mathbb{R}^{n\times n}$ can be defined as

$$\frac{\partial f(X)}{\partial X} \circ \Delta = \lim_{h\to 0} \frac{f(X + h\Delta) - f(X)}{h}$$

Important examples:

- $\frac{\partial \mathrm{Tr}(AX)}{\partial X} \circ \Delta = \mathrm{Tr}(A\Delta)$.
- $\frac{\partial X^{-1}}{\partial X} \circ \Delta = -X^{-1}\Delta X^{-1}$.
- $\frac{\partial \mathrm{Tr}(AXBX^TC)}{\partial X} \circ \Delta = \mathrm{Tr}(A\Delta BX^TC + AXB\Delta^TC) = \mathrm{Tr}(\left[BX^TCA + B^TX^TC^TA^T\right]\Delta)$.

# Some other useful results from matrix analysis

The derivative of a function $f : \mathbb{R}^{n\times n} \to D$ in the direction $\Delta \in \mathbb{R}^{n\times n}$ can be defined as

$$\frac{\partial f(X)}{\partial X} \circ \Delta = \lim_{h\to 0} \frac{f(X+h\Delta) - f(X)}{h}$$

Important examples:

- $\frac{\partial \mathrm{Tr}(AX)}{\partial X} \circ \Delta = \mathrm{Tr}(A\Delta)$.
- $\frac{\partial X^{-1}}{\partial X} \circ \Delta = -X^{-1}\Delta X^{-1}$.
- $\frac{\partial \mathrm{Tr}(AXBX^TC)}{\partial X} \circ \Delta = \mathrm{Tr}(A\Delta BX^TC + AXB\Delta^TC) = \mathrm{Tr}(\left[BX^TCA + B^TX^TC^TA^T\right]\Delta)$.

# Contents

# Exploiting Gradient Information

When implementing Newton type methods of the form

$$x = x^- - (M^-)^{-1} F'(x^-)^T \ , \quad x^+ = x - M^{-1} F'(x)^T \ , \quad \text{and so on}$$

we have to compute the gradient $F'(x^-)$ at the previous iterate and the gradient $F'(x)^T$ at the current iterate.

Since we evaluate the gradient at two points anyhow, we can obtain the directional estimate

$$F''(x)(x - x^-) \approx F'(x)^T - F'(x^-)^T$$

Can we use this relation to improve our next Hessian approximation $M^+ \approx F''(x^+)$?

# Exploiting Gradient Information

When implementing Newton type methods of the form

$$x = x^- - (M^-)^{-1} F'(x^-)^T \ , \quad x^+ = x - M^{-1} F'(x)^T \ , \quad \text{and so on}$$

we have to compute the gradient $F'(x^-)$ at the previous iterate and the gradient $F'(x)^T$ at the current iterate.

Since we evaluate the gradient at two points anyhow, we can obtain the directional estimate

$$F''(x)(x - x^-) \approx F'(x)^T - F'(x^-)^T$$

Can we use this relation to improve our next Hessian approximation $M^+ \approx F''(x^+)$?

# Exploiting Gradient Information

Let $M$ be our current Hessian approximation. The relation

$$F''(x)d \approx y \quad \text{with} \quad d = x - x^- \quad \text{and} \quad y = F'(x)^T - F'(x^-)^T$$

motivates to improve our current estimate of $F''$ constructing $M^+$ by solving

$$\min_{M^+} \frac{1}{2} \left\| M^+ - M \right\|^2 \quad \text{s.t.} \quad M^+ d = y$$

for a suitable matrix norm $\|\cdot\|$.

# Exploiting Gradient Information

If we work with Frobenius norms, we can solve the optimization problem

$$\min_{M^+} \frac{1}{2} \left\| M^+ - M \right\|_F^2 \quad \text{s.t.} \quad M^+ d = y$$

explicitly. Here, the Frobenius norm is given by

$$\|X\|_F^2 = \mathrm{Tr}(XX^T) \ .$$

For this aim, we work out the optimality conditions

$$0 = (M^+ - M)^T + d\lambda^T \quad \text{and} \quad M^+ d = y \ .$$

# Exploiting Gradient Information

If we work with Frobenius norms, we can solve the optimization problem

$$\min_{M^+} \frac{1}{2} \left\| M^+ - M \right\|_F^2 \quad \text{s.t.} \quad M^+ d = y$$

explicitly. Here, the Frobenius norm is given by

$$\|X\|_F^2 = \mathrm{Tr}(XX^T) \; .$$

For this aim, we work out the optimality conditions

$$0 = (M^+ - M)^T + d\lambda^T \quad \text{and} \quad M^+ d = y \; .$$

# Broyden's update formula

The multiplier $\lambda$ can be found by eliminating $M^+$ from the stationarity condition,

$$M^+ = M - \lambda d^T \ ,$$

and substituting into the directional equality constraint,

$$M^+ d = (M - \lambda d^T)d = y$$

which yields $\lambda = \frac{1}{d^T d}(Md - y)$. The corresponding update formula,

$$M^+ = M - \frac{(Md - y)d^T}{d^T d}$$

is called Broyden's matrix update.

# Broyden's update formula

The multiplier $\lambda$ can be found by eliminating $M^+$ from the stationarity condition,

$$M^+ = M - \lambda d^T \ ,$$

and substituting into the directional equality constraint,

$$M^+ d = (M - \lambda d^T) d = y$$

which yields $\lambda = \frac{1}{d^T d}(Md - y)$. The corresponding update formula,

$$M^+ = M - \frac{(Md - y) d^T}{d^T d}$$

is called Broyden's matrix update.

# Inverse Broyden's update formula

Broyden's updates turns out to be a rank-1 update,

$$M^{+} = M - \frac{(Md - y)d^{T}}{d^{T}d} \; .$$

Assuming that we have already computed $M^{-1}$, Woodbury's matrix inversion formula yields a direct update of the inverse:

$$\begin{aligned} (M^{+})^{-1} \quad &= \quad M^{-1} + M^{-1}\frac{Md - y}{d^{T}d}\left(1 - d^{T}M^{-1}\frac{Md - y}{d^{T}d}\right)^{-1} d^{T}M^{-1} \\ &= \quad M^{-1} + \frac{(d - M^{-1}y)d^{T}M^{-1}}{d^{T}M^{-1}y} \; . \end{aligned}$$

# Inverse Broyden's update formula

Broyden's update formula

$$(M^+)^{-1} = M^{-1} + \frac{(d - M^{-1}y)d^T M^{-1}}{d^T M^{-1} y} \ .$$

solves two problems at the same time:

- we don't need to compute any second order derivatives
- we can directly compute $(M^+)^{-1}$, no inversion needed.

But: $M^+$ may be non-symmetric even if the original matrix $M$ was symmetric.

# Inverse Broyden's update formula

Broyden's update formula

$$(M^+)^{-1} = M^{-1} + \frac{(d - M^{-1}y)d^T M^{-1}}{d^T M^{-1} y} \ .$$

solves two problems at the same time:

- we don't need to compute any second order derivatives
- we can directly compute $(M^+)^{-1}$, no inversion needed.

But: $M^+$ may be non-symmetric even if the original matrix $M$ was symmetric.

# Contents

# Broyden-Fletcher-Goldfarb-Shanno Updates

Broyden, Fletcher, Goldfarb, and Shanno suggested a technique to improve Broyden's original update formula. The idea is to maintain the symmetry of the updates by solving

$$\min_{M^+} \frac{1}{2} \left\| M^+ - M \right\|^2 \quad \text{s.t.} \quad \begin{cases} (M^+)^T d = y \\ M^+ d = y \ . \end{cases}$$

Here, the norm is (mainly for computational reasons) weighted in very particular way (assume $M$ is positive definite):

$$\left\| M^+ - M \right\|^2 = \mathrm{Tr} \left( W^{\frac{1}{2}} (M^+ - M)^T M^{-1} (M^+ - M) W^{\frac{1}{2}} \right) ,$$

where $W^{\frac{1}{2}}$ can be any symmetric positive definite weighting matrix satisfying $Wy = d$.

# Broyden-Fletcher-Goldfarb-Shanno Updates

Broyden, Fletcher, Goldfarb, and Shanno suggested a technique to improve Broyden's original update formula. The idea is to maintain the symmetry of the updates by solving

$$\min_{M^+} \frac{1}{2} \left\| M^+ - M \right\|^2 \quad \text{s.t.} \quad \begin{cases} (M^+)^T d = y \\ M^+ d = y \ . \end{cases}$$

Here, the norm is (mainly for computational reasons) weighted in very particular way (assume $M$ is positive definite):

$$\left\| M^+ - M \right\|^2 = \mathrm{Tr} \left( W^{\frac{1}{2}} (M^+ - M)^T M^{-1} (M^+ - M) W^{\frac{1}{2}} \right) \ ,$$

where $W^{\frac{1}{2}}$ can be any symmetric positive definite weighting matrix satisfying $Wy = d$.

# Broyden-Fletcher-Goldfarb-Shanno Updates

The first order necessary (and sufficient) optimality conditions take the form

$$0 = W(M^+ - M)^T M^{-1} + d\lambda^T + \mu d^T \quad \text{and} \quad \begin{cases} (M^+)^T d = y \\ M^+ d = y \end{cases}$$

Here, we assume $M = M^T$. It is easy to check that these conditions are satisfied for the symmetric rank-2 update

$$M^+ = M + \frac{yy^T}{y^T d} - \frac{Mdd^T M}{d^T Md} .$$

This is called the BFGS update formula; symmetry is maintained.

# Broyden-Fletcher-Goldfarb-Shanno Updates

The first order necessary (and sufficient) optimality conditions take the form

$$0 = W(M^+ - M)^T M^{-1} + d\lambda^T + \mu d^T \quad \text{and} \quad \begin{cases} (M^+)^T d = y \\ M^+ d = y \end{cases}$$

Here, we assume $M = M^T$. It is easy to check that these conditions are satisfied for the symmetric rank-2 update

$$M^+ = M + \frac{yy^T}{y^T d} - \frac{Mdd^T M}{d^T Md} \; .$$

This is called the BFGS update formula; symmetry is maintained.

# Broyden-Fletcher-Goldfarb-Shanno Updates

Similar to Broyden updates the BFGS update can be applied through Woodbury's formula. This yields a direct update for the inverse of $M$, which has the form

$$(M^+)^{-1} = \left(I - \frac{dy^T}{d^Ty}\right) M^{-1} \left(I - \frac{dy^T}{d^Ty}\right)^T + \frac{dd^T}{d^Ty} \ .$$

Notice that if $F$ is strictly convex, the term

$$d^T y = \left(x - x^-\right)^T \left(F'(x)^T - F'(x^-)^T\right) \approx \left(x - x^-\right)^T F''(x) \left(x - x^-\right)$$

can be expected to be positive. (there are many variants of BFGS around; some additionally maintain the positive definiteness of $M$; others work with "limited memory")

# Broyden-Fletcher-Goldfarb-Shanno Updates

Similar to Broyden updates the BFGS update can be applied through Woodbury's formula. This yields a direct update for the inverse of $M$, which has the form

$$(M^+)^{-1} = \left(I - \frac{dy^T}{d^T y}\right) M^{-1} \left(I - \frac{dy^T}{d^T y}\right)^T + \frac{dd^T}{d^T y} \; .$$

Notice that if $F$ is strictly convex, the term

$$d^T y = \left(x - x^-\right)^T \left(F'(x)^T - F'(x^-)^T\right) \approx \left(x - x^-\right)^T F''(x) \left(x - x^-\right)$$

can be expected to be positive. (there are many variants of BFGS around; some additionally maintain the positive definiteness of $M$; others work with “limited memory”)

# Broyden-Fletcher-Goldfarb-Shanno Updates

Similar to Broyden updates the BFGS update can be applied through Woodbury's formula. This yields a direct update for the inverse of $M$, which has the form

$$(M^+)^{-1} = \left(I - \frac{dy^T}{d^T y}\right) M^{-1} \left(I - \frac{dy^T}{d^T y}\right)^T + \frac{dd^T}{d^T y} \ .$$

Notice that if $F$ is strictly convex, the term

$$d^T y = \left(x - x^-\right)^T \left(F'(x)^T - F'(x^-)^T\right) \approx \left(x - x^-\right)^T F''(x) \left(x - x^-\right)$$

can be expected to be positive. (there are many variants of BFGS around; some additionally maintain the positive definiteness of $M$; others work with "limited memory")