Numerical Optimal Control
Lecture 3: Newton method & SQP

Sébastien Gros

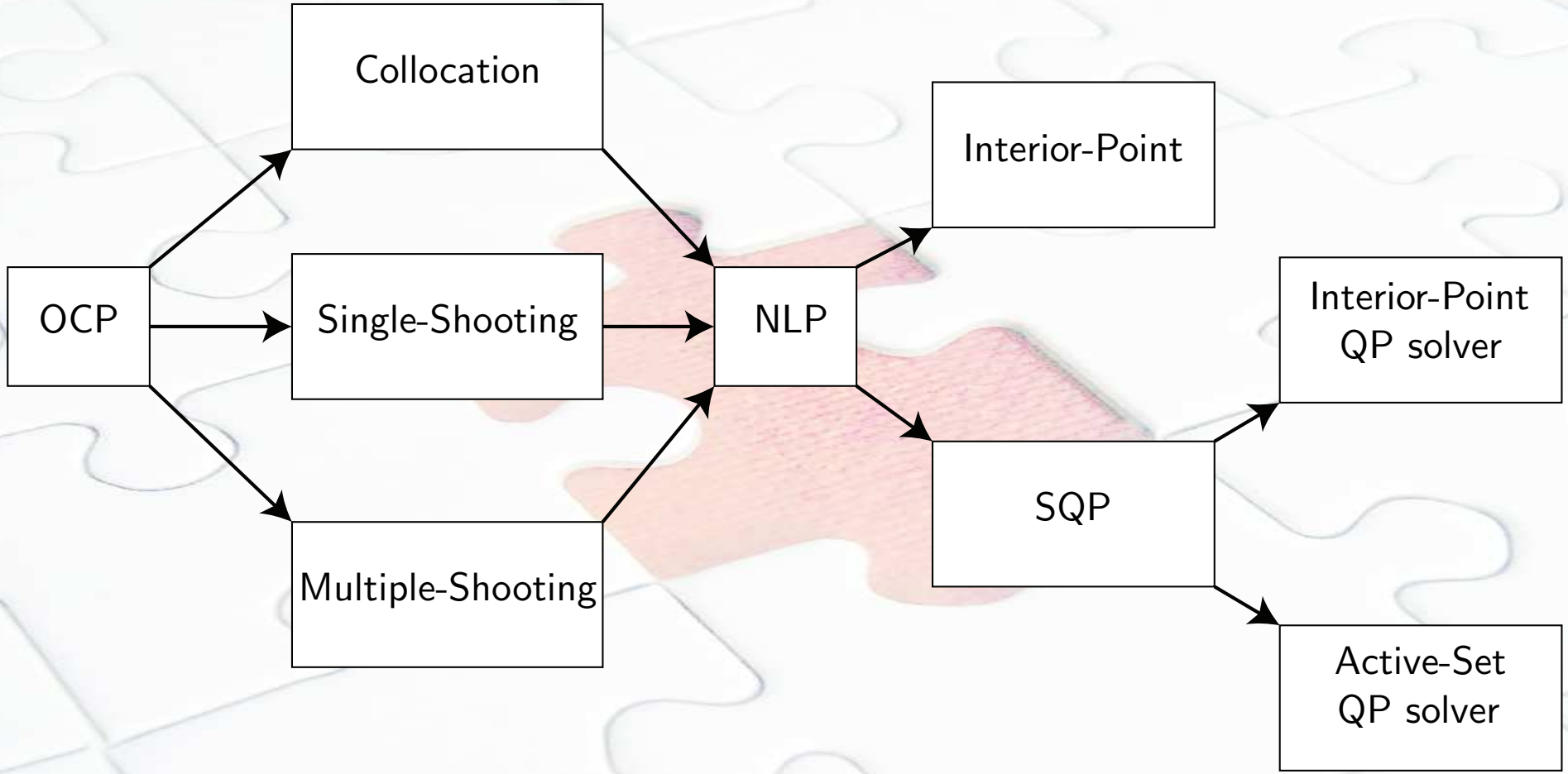
ITK NTNU

NTNU PhD course

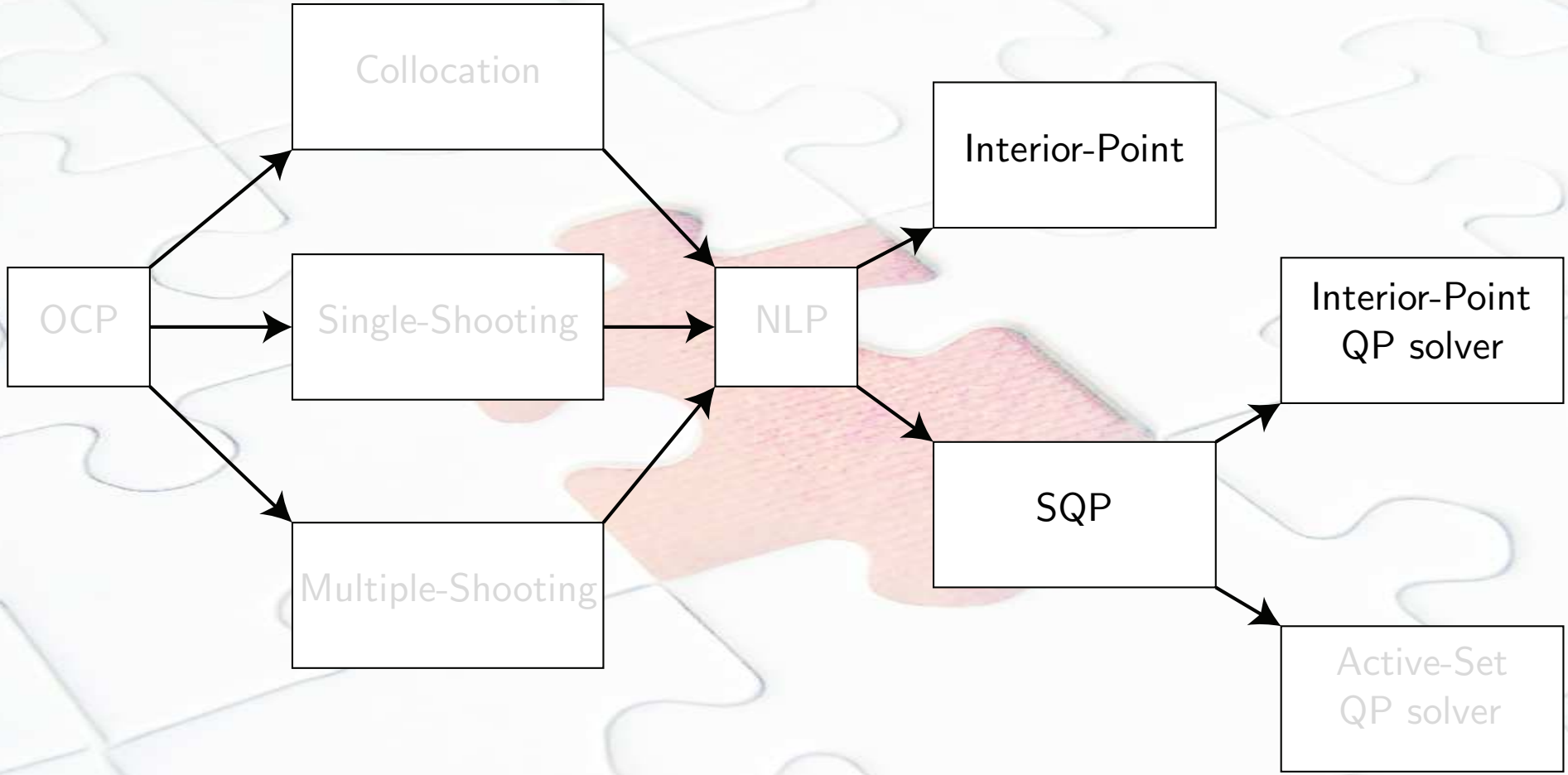
Outline

1. The Newton method
2. Newton on the KKT conditions
3. The reduced Newton step (unconstrained problems)
4. The merit function - Line-search for constrained problems
5. Newton-type methods
6. Sequential Quadratic Programming

Survival map of Direct Optimal Control

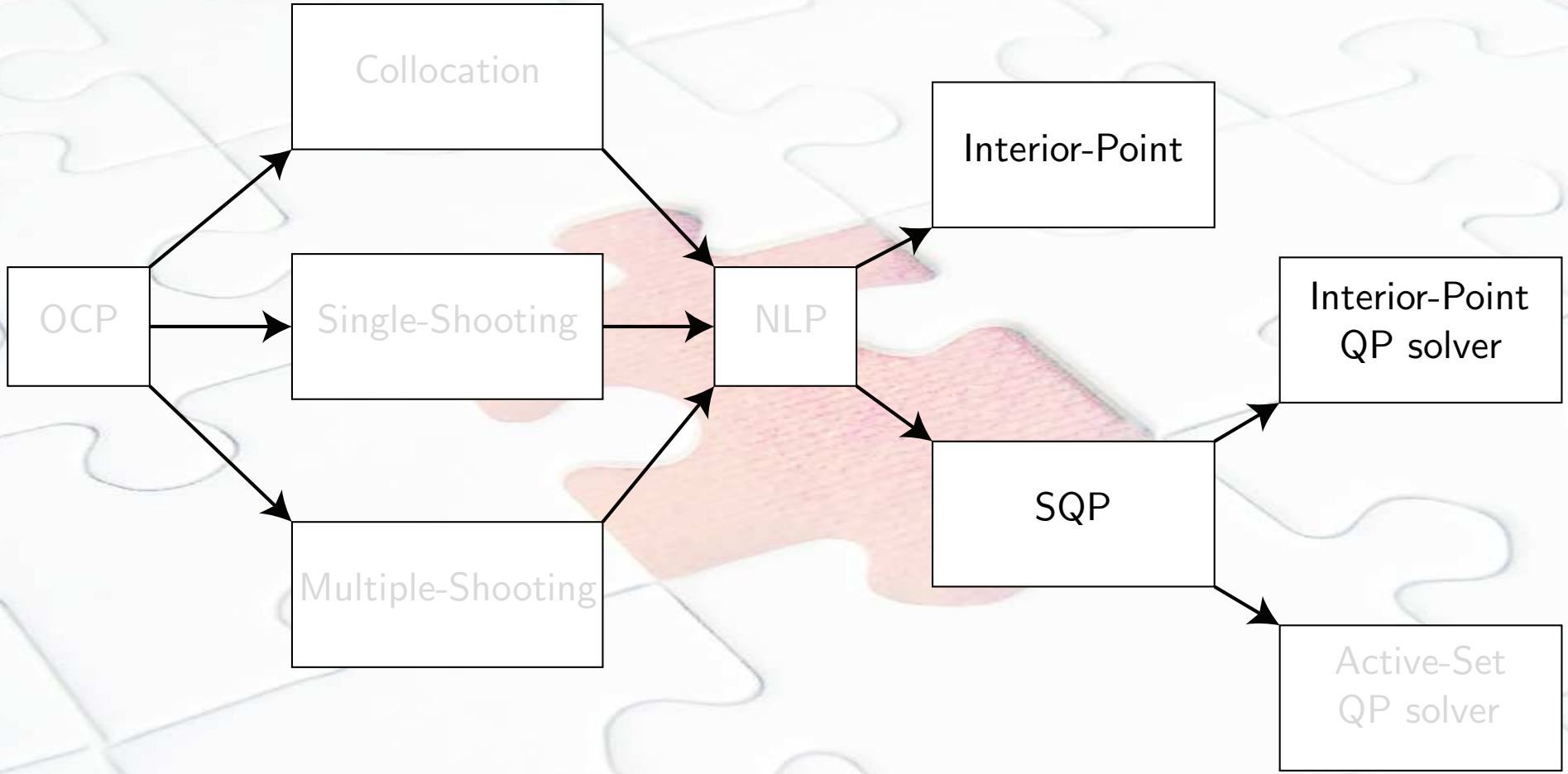


Survival map of Direct Optimal Control



Newton - a general-purpose sledgehammer for algebraic equations...

Survival map of Direct Optimal Control



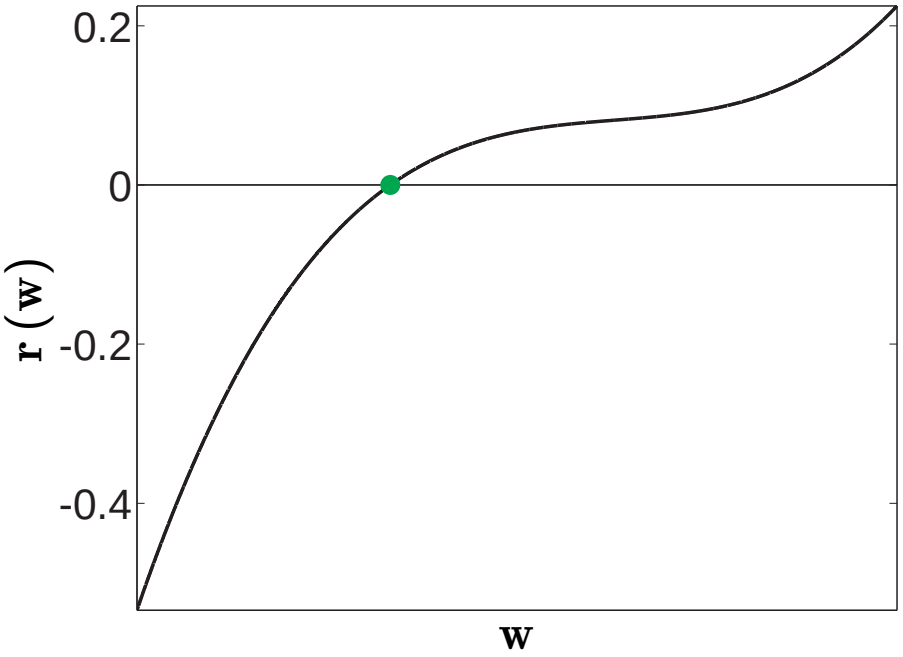
Newton - a general-purpose sledgehammer for algebraic equations...
... will be used to solve the KKT conditions !!

Outline

1. The Newton method
2. Newton on the KKT conditions
3. The reduced Newton step (unconstrained problems)
4. The merit function - Line-search for constrained problems
5. Newton-type methods
6. Sequential Quadratic Programming

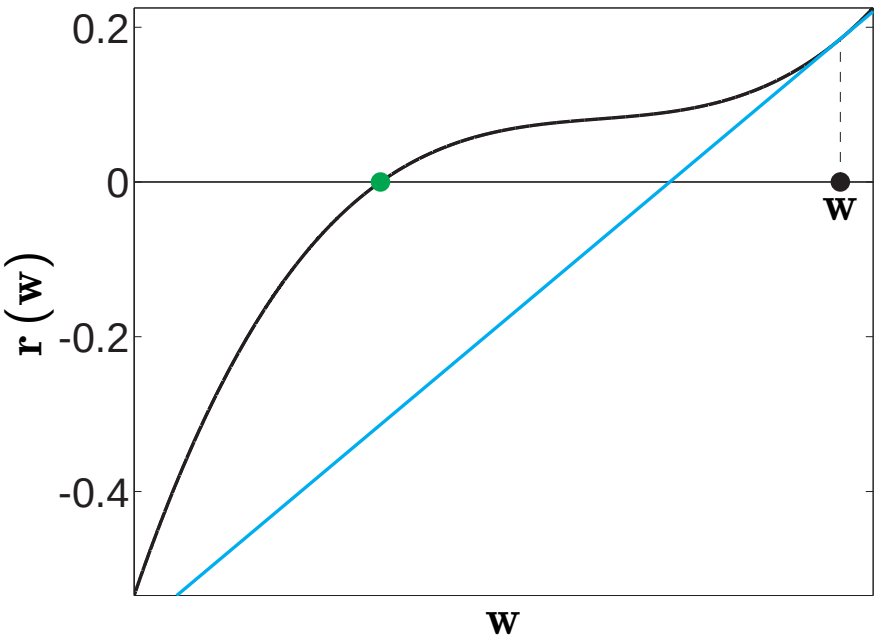
Core idea

Goal: solve $\mathbf{r}(\mathbf{w}) = 0$... **how ?!?**



Core idea

Goal: solve $\mathbf{r}(\mathbf{w}) = 0$... **how ?!?**

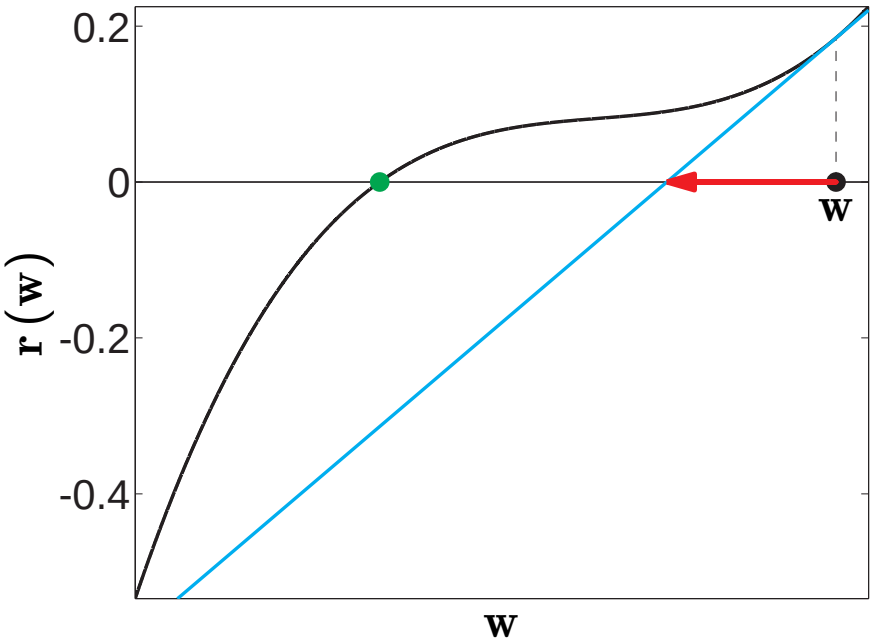


Key idea: guess $\mathbf{w}$, iterate the linear model:

$$\mathbf{r}(\mathbf{w} + \Delta\mathbf{w}) \approx \mathbf{r}(\mathbf{w}) + \nabla\mathbf{r}(\mathbf{w})^{\top} \Delta\mathbf{w} = 0$$

Core idea

Goal: solve $\mathbf{r}(\mathbf{w}) = 0$... **how ?!?**

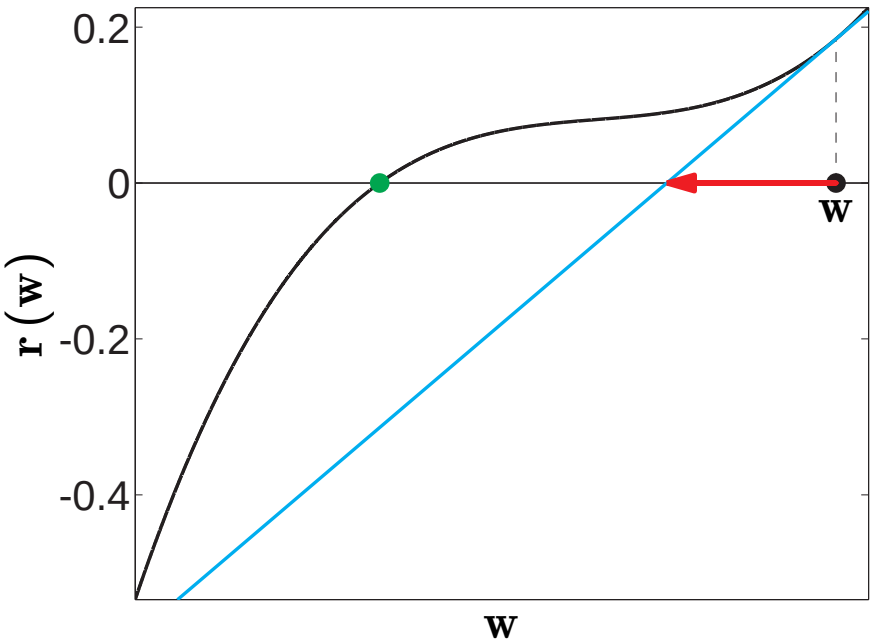


Key idea: guess $\mathbf{w}$, iterate the linear model:

$$\mathbf{r}(\mathbf{w} + \Delta\mathbf{w}) \approx \mathbf{r}(\mathbf{w}) + \nabla\mathbf{r}(\mathbf{w})^{\top} \Delta\mathbf{w} = 0$$

Core idea

Goal: solve $\mathbf{r}(\mathbf{w}) = 0$... **how ?!?**



Key idea: guess $\mathbf{w}$, iterate the linear model:

$$\mathbf{r}(\mathbf{w} + \Delta\mathbf{w}) \approx \mathbf{r}(\mathbf{w}) + \nabla\mathbf{r}(\mathbf{w})^{\top} \Delta\mathbf{w} = 0$$

Algorithm: Newton method

Input: $\mathbf{w}$, Tol

while $\|\mathbf{r}(\mathbf{w})\| \geq \text{tol}$ **do**

Compute

$$\mathbf{r}(\mathbf{w}) \quad \text{and} \quad \nabla\mathbf{r}(\mathbf{w})$$

Compute the Newton direction

$$\nabla\mathbf{r}(\mathbf{w})^{\mathsf{T}} \Delta\mathbf{w} = -\mathbf{r}(\mathbf{w})$$

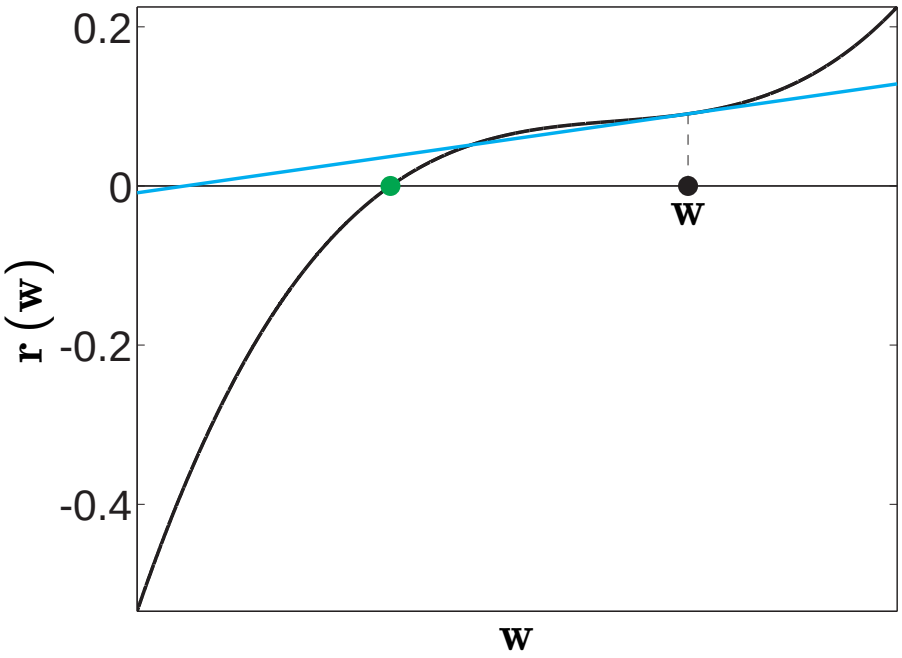
Newton step

$$\mathbf{w} \leftarrow \mathbf{w} + \Delta\mathbf{w}$$

return $\mathbf{w}$

Core idea

Goal: solve $\mathbf{r}(\mathbf{w}) = 0$... **how ?!?**



Key idea: guess $\mathbf{w}$, iterate the linear model:

$$\mathbf{r}(\mathbf{w} + \Delta\mathbf{w}) \approx \mathbf{r}(\mathbf{w}) + \nabla\mathbf{r}(\mathbf{w})^{\top} \Delta\mathbf{w} = 0$$

Algorithm: Newton method

Input: $\mathbf{w}$, Tol

while $\|\mathbf{r}(\mathbf{w})\| \geq \text{tol}$ **do**

Compute

$$\mathbf{r}(\mathbf{w}) \quad \text{and} \quad \nabla\mathbf{r}(\mathbf{w})$$

Compute the Newton direction

$$\nabla\mathbf{r}(\mathbf{w})^{\mathsf{T}} \Delta\mathbf{w} = -\mathbf{r}(\mathbf{w})$$

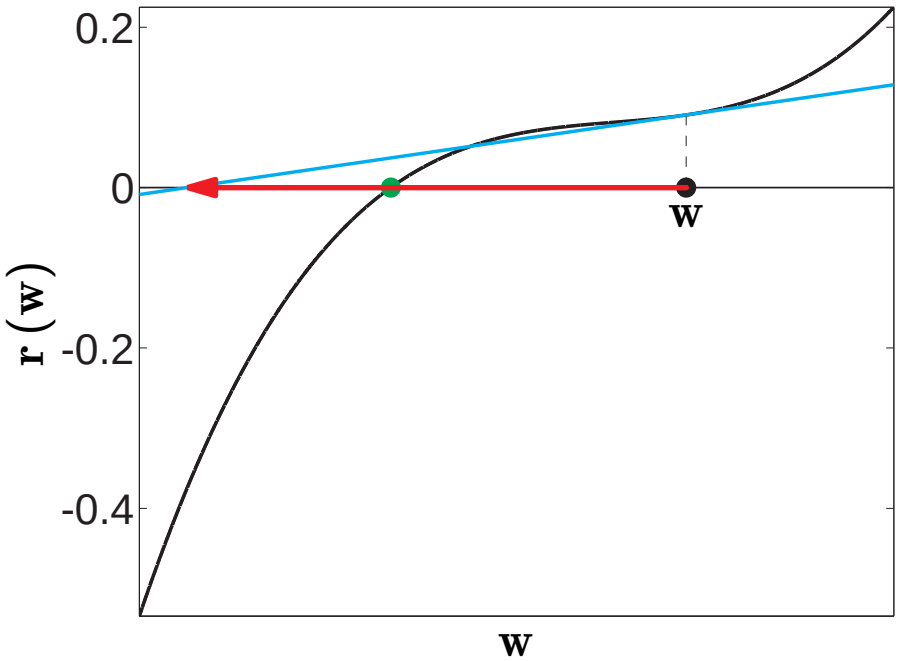
Newton step

$$\mathbf{w} \leftarrow \mathbf{w} + \Delta\mathbf{w}$$

return $\mathbf{w}$

Core idea

Goal: solve $\mathbf{r}(\mathbf{w}) = 0$... **how ?!?**



Key idea: guess $\mathbf{w}$, iterate the linear model:

$$\mathbf{r}(\mathbf{w} + \Delta\mathbf{w}) \approx \mathbf{r}(\mathbf{w}) + \nabla\mathbf{r}(\mathbf{w})^{\top} \Delta\mathbf{w} = 0$$

Algorithm: Newton method

Input: $\mathbf{w}$, Tol

while $\|\mathbf{r}(\mathbf{w})\| \geq \text{tol}$ **do**

Compute

$$\mathbf{r}(\mathbf{w}) \quad \text{and} \quad \nabla\mathbf{r}(\mathbf{w})$$

Compute the Newton direction

$$\nabla\mathbf{r}(\mathbf{w})^{\mathsf{T}} \Delta\mathbf{w} = -\mathbf{r}(\mathbf{w})$$

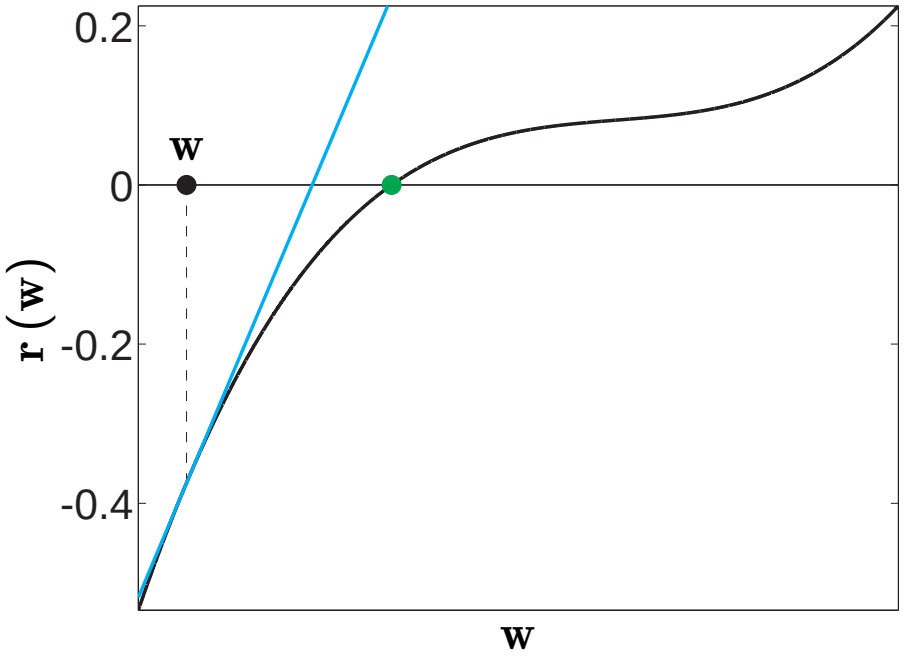
Newton step

$$\mathbf{w} \leftarrow \mathbf{w} + \Delta\mathbf{w}$$

return $\mathbf{w}$

Core idea

Goal: solve $\mathbf{r}(\mathbf{w}) = 0$... **how ?!?**



Key idea: guess $\mathbf{w}$, iterate the linear model:

$$\mathbf{r}(\mathbf{w} + \Delta\mathbf{w}) \approx \mathbf{r}(\mathbf{w}) + \nabla\mathbf{r}(\mathbf{w})^{\top} \Delta\mathbf{w} = 0$$

Algorithm: Newton method

Input: $\mathbf{w}$, Tol

while $\|\mathbf{r}(\mathbf{w})\| \geq \text{tol}$ **do**

Compute

$$\mathbf{r}(\mathbf{w}) \quad \text{and} \quad \nabla\mathbf{r}(\mathbf{w})$$

Compute the Newton direction

$$\nabla\mathbf{r}(\mathbf{w})^{\mathsf{T}} \Delta\mathbf{w} = -\mathbf{r}(\mathbf{w})$$

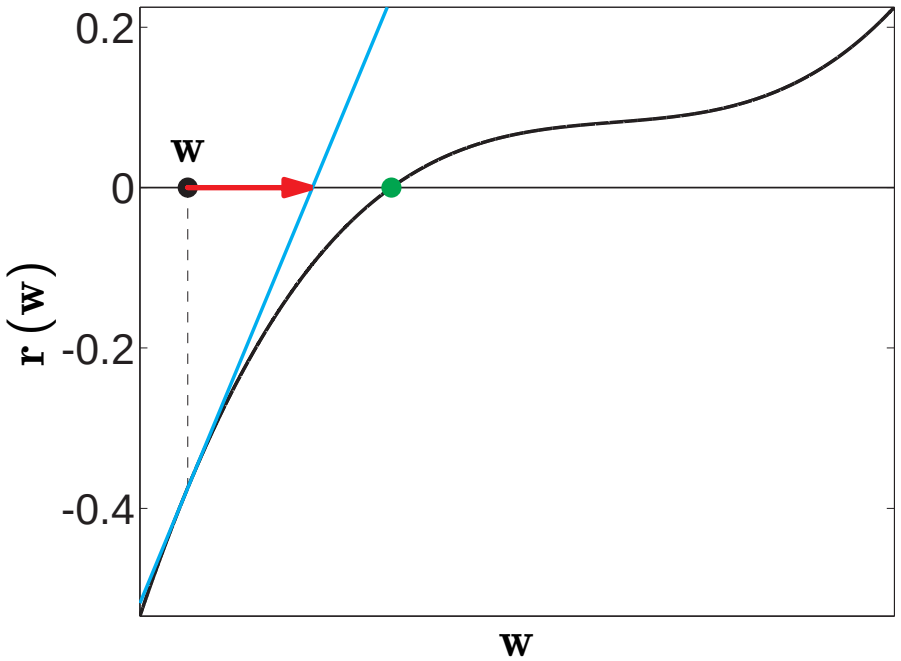
Newton step

$$\mathbf{w} \leftarrow \mathbf{w} + \Delta\mathbf{w}$$

return $\mathbf{w}$

Core idea

Goal: solve $\mathbf{r}(\mathbf{w}) = 0$... **how ?!?**



Key idea: guess $\mathbf{w}$, iterate the linear model:

$$\mathbf{r}(\mathbf{w} + \Delta\mathbf{w}) \approx \mathbf{r}(\mathbf{w}) + \nabla\mathbf{r}(\mathbf{w})^{\top} \Delta\mathbf{w} = 0$$

Algorithm: Newton method

Input: $\mathbf{w}$, Tol

while $\|\mathbf{r}(\mathbf{w})\| \geq \text{tol}$ **do**

Compute

$$\mathbf{r}(\mathbf{w}) \quad \text{and} \quad \nabla\mathbf{r}(\mathbf{w})$$

Compute the Newton direction

$$\nabla\mathbf{r}(\mathbf{w})^{\mathsf{T}} \Delta\mathbf{w} = -\mathbf{r}(\mathbf{w})$$

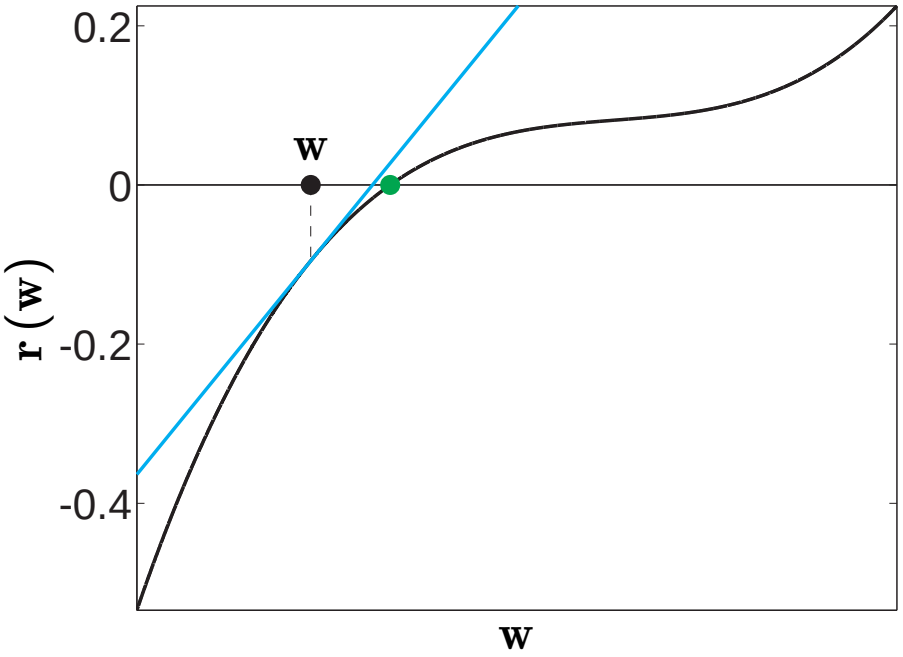
Newton step

$$\mathbf{w} \leftarrow \mathbf{w} + \Delta\mathbf{w}$$

return $\mathbf{w}$

Core idea

Goal: solve $\mathbf{r}(\mathbf{w}) = 0$... **how ?!?**



Key idea: guess $\mathbf{w}$, iterate the linear model:

$$\mathbf{r}(\mathbf{w} + \Delta\mathbf{w}) \approx \mathbf{r}(\mathbf{w}) + \nabla\mathbf{r}(\mathbf{w})^\top \Delta\mathbf{w} = 0$$

Algorithm: Newton method

Input: $\mathbf{w}$, Tol

while $\|\mathbf{r}(\mathbf{w})\| \geq \text{tol}$ **do**

Compute

$$\mathbf{r}(\mathbf{w}) \quad \text{and} \quad \nabla\mathbf{r}(\mathbf{w})$$

Compute the Newton direction

$$\nabla\mathbf{r}(\mathbf{w})^\mathsf{T} \Delta\mathbf{w} = -\mathbf{r}(\mathbf{w})$$

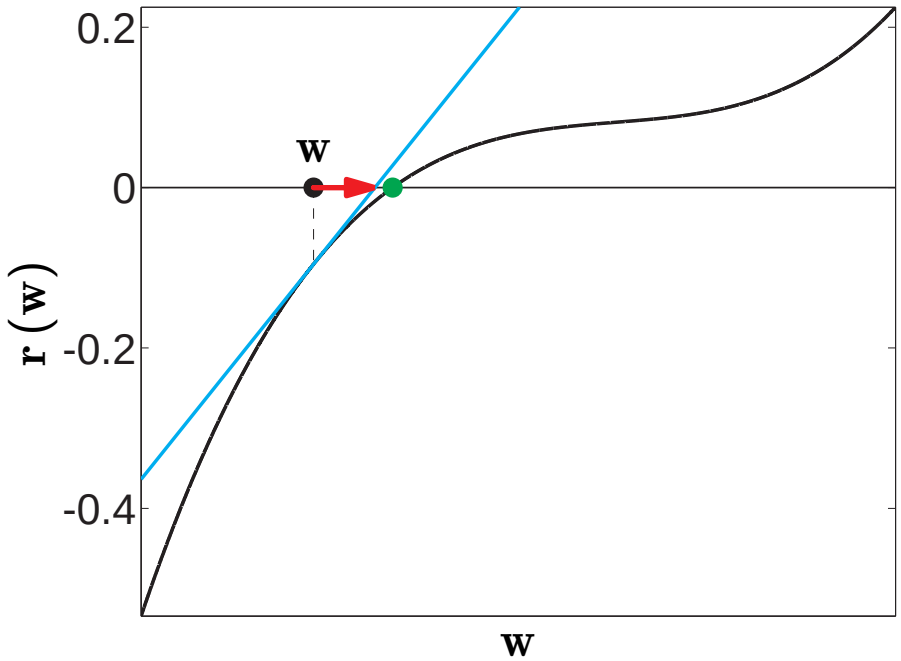
Newton step

$$\mathbf{w} \leftarrow \mathbf{w} + \Delta\mathbf{w}$$

return $\mathbf{w}$

Core idea

Goal: solve $\mathbf{r}(\mathbf{w}) = 0$... **how ?!?**



Key idea: guess $\mathbf{w}$, iterate the linear model:

$$\mathbf{r}(\mathbf{w} + \Delta\mathbf{w}) \approx \mathbf{r}(\mathbf{w}) + \nabla\mathbf{r}(\mathbf{w})^{\top} \Delta\mathbf{w} = 0$$

Algorithm: Newton method

Input: $\mathbf{w}$, Tol

while $\|\mathbf{r}(\mathbf{w})\| \geq \text{tol}$ **do**

Compute

$$\mathbf{r}(\mathbf{w}) \quad \text{and} \quad \nabla\mathbf{r}(\mathbf{w})$$

Compute the Newton direction

$$\nabla\mathbf{r}(\mathbf{w})^{\mathsf{T}} \Delta\mathbf{w} = -\mathbf{r}(\mathbf{w})$$

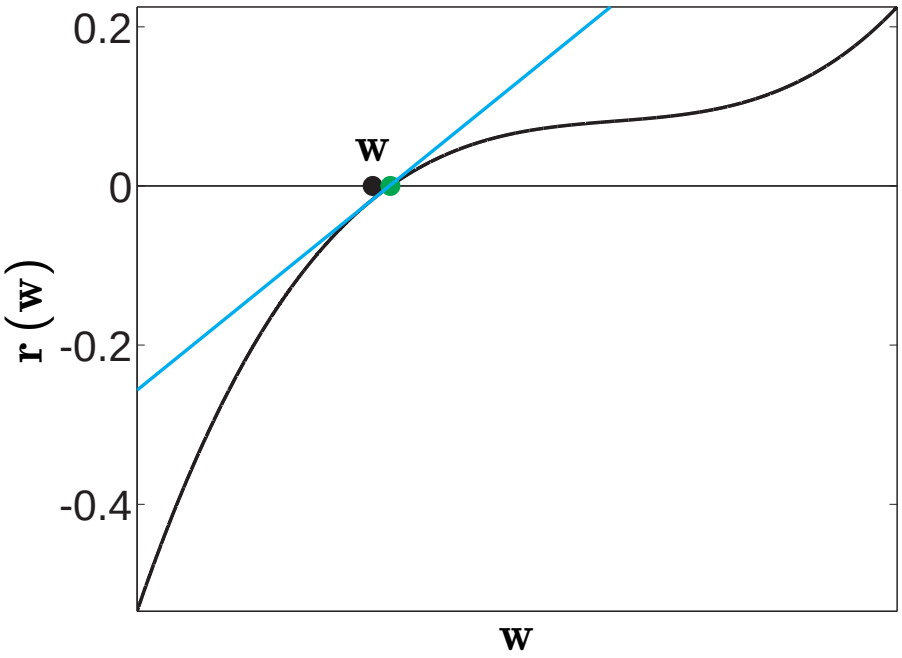
Newton step

$$\mathbf{w} \leftarrow \mathbf{w} + \Delta\mathbf{w}$$

return $\mathbf{w}$

Core idea

Goal: solve $\mathbf{r}(\mathbf{w}) = 0$... **how ?!?**



Key idea: guess $\mathbf{w}$, iterate the linear model:

$$\mathbf{r}(\mathbf{w} + \Delta\mathbf{w}) \approx \mathbf{r}(\mathbf{w}) + \nabla\mathbf{r}(\mathbf{w})^{\top} \Delta\mathbf{w} = 0$$

Algorithm: Newton method

Input: $\mathbf{w}$, Tol

while $\|\mathbf{r}(\mathbf{w})\| \geq \text{tol}$ **do**

Compute

$$\mathbf{r}(\mathbf{w}) \quad \text{and} \quad \nabla\mathbf{r}(\mathbf{w})$$

Compute the Newton direction

$$\nabla\mathbf{r}(\mathbf{w})^{\mathsf{T}} \Delta\mathbf{w} = -\mathbf{r}(\mathbf{w})$$

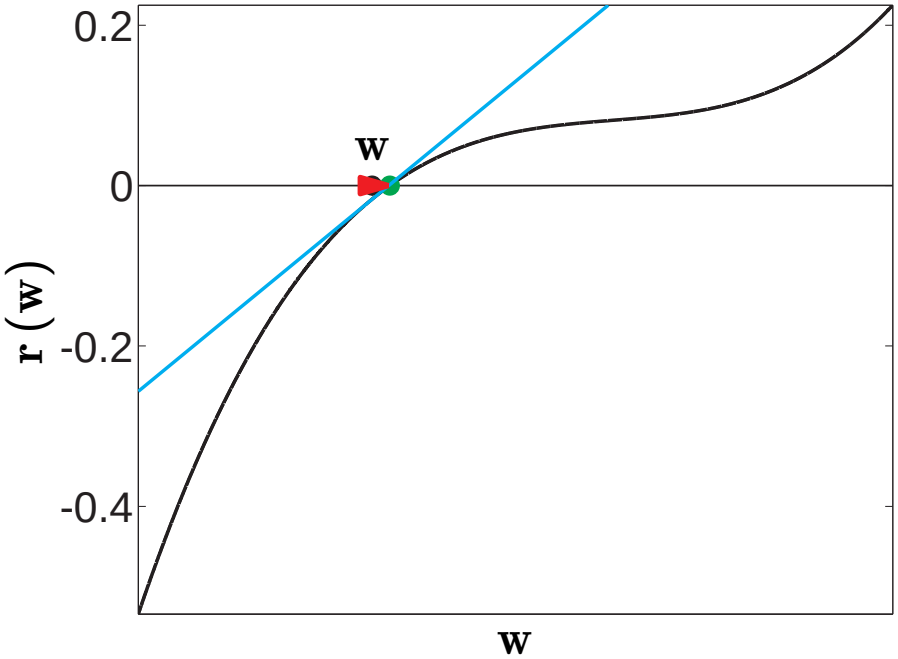
Newton step

$$\mathbf{w} \leftarrow \mathbf{w} + \Delta\mathbf{w}$$

return $\mathbf{w}$

Core idea

Goal: solve $\mathbf{r}(\mathbf{w}) = 0$... **how ?!?**



Key idea: guess $\mathbf{w}$, iterate the linear model:

$$\mathbf{r}(\mathbf{w} + \Delta\mathbf{w}) \approx \mathbf{r}(\mathbf{w}) + \nabla\mathbf{r}(\mathbf{w})^{\top} \Delta\mathbf{w} = 0$$

Algorithm: Newton method

Input: $\mathbf{w}$, Tol

while $\|\mathbf{r}(\mathbf{w})\| \geq \text{tol}$ **do**

Compute

$$\mathbf{r}(\mathbf{w}) \quad \text{and} \quad \nabla\mathbf{r}(\mathbf{w})$$

Compute the Newton direction

$$\nabla\mathbf{r}(\mathbf{w})^{\mathsf{T}} \Delta\mathbf{w} = -\mathbf{r}(\mathbf{w})$$

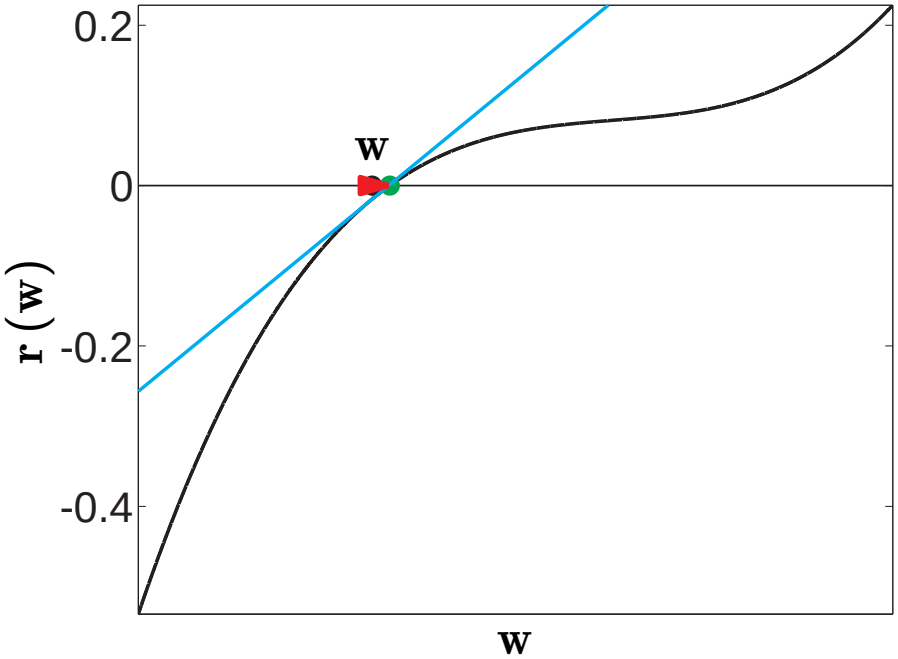
Newton step

$$\mathbf{w} \leftarrow \mathbf{w} + \Delta\mathbf{w}$$

return $\mathbf{w}$

Core idea

Goal: solve $\mathbf{r}(\mathbf{w}) = 0$... **how ?!?**



Key idea: guess $\mathbf{w}$, iterate the linear model:

$$\mathbf{r}(\mathbf{w} + \Delta\mathbf{w}) \approx \mathbf{r}(\mathbf{w}) + \nabla\mathbf{r}(\mathbf{w})^{\top} \Delta\mathbf{w} = 0$$

Algorithm: Newton method

Input: $\mathbf{w}$, Tol

while $\|\mathbf{r}(\mathbf{w})\| \geq \text{tol}$ **do**

Compute

$$\mathbf{r}(\mathbf{w}) \quad \text{and} \quad \nabla\mathbf{r}(\mathbf{w})$$

Compute the Newton direction

$$\nabla\mathbf{r}(\mathbf{w})^{\top} \Delta\mathbf{w} = -\mathbf{r}(\mathbf{w})$$

Newton step

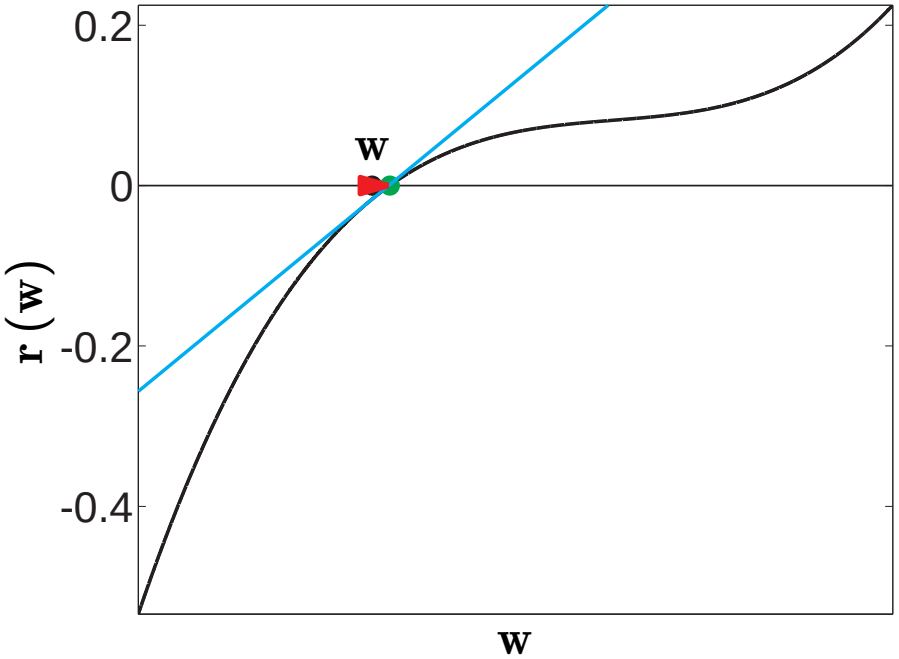
$$\mathbf{w} \leftarrow \mathbf{w} + \Delta\mathbf{w}$$

return $\mathbf{w}$

- This is a **full-step** Newton iteration

Core idea

Goal: solve $\mathbf{r}(\mathbf{w}) = 0$... **how ?!?**



Key idea: guess $\mathbf{w}$, iterate the linear model:

$$\mathbf{r}(\mathbf{w} + \Delta\mathbf{w}) \approx \mathbf{r}(\mathbf{w}) + \nabla\mathbf{r}(\mathbf{w})^{\top} \Delta\mathbf{w} = 0$$

Algorithm: Newton method

Input: $\mathbf{w}$, Tol

while $\|\mathbf{r}(\mathbf{w})\| \geq \text{tol}$ **do**

Compute

$$\mathbf{r}(\mathbf{w}) \quad \text{and} \quad \nabla\mathbf{r}(\mathbf{w})$$

Compute the Newton direction

$$\nabla\mathbf{r}(\mathbf{w})^{\top} \Delta\mathbf{w} = -\mathbf{r}(\mathbf{w})$$

Newton step, $t \in]0, 1]$

$$\mathbf{w} \leftarrow \mathbf{w} + t\Delta\mathbf{w}$$

return $\mathbf{w}$

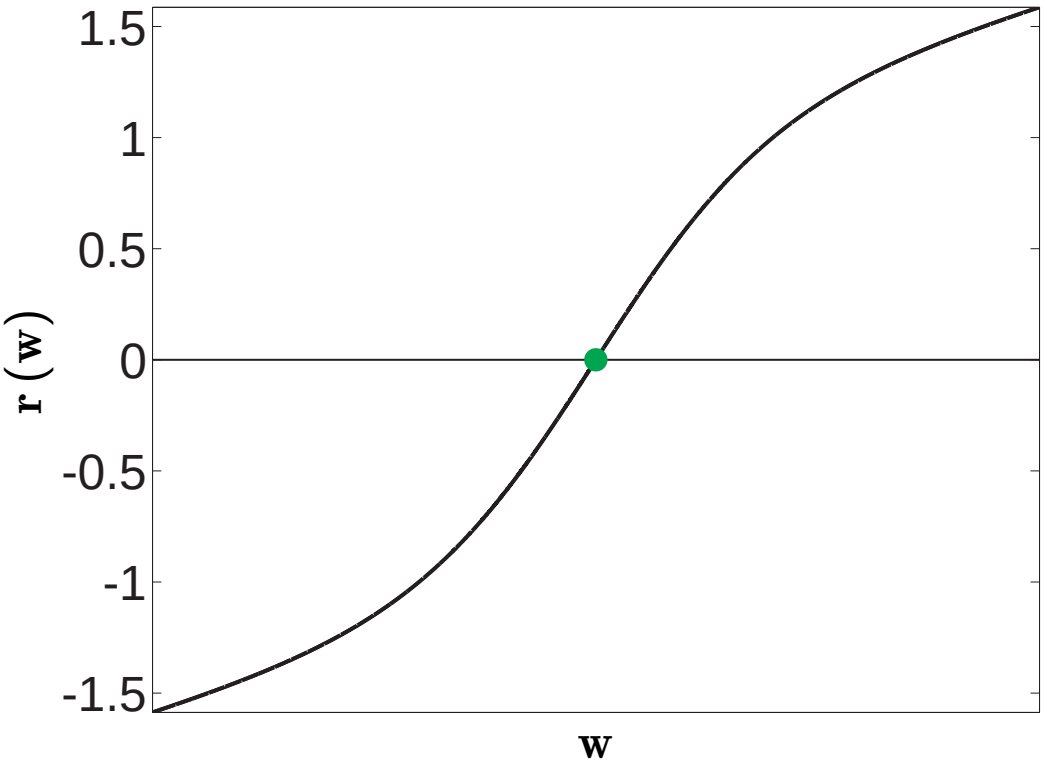
- This is a **full-step** Newton iteration
- Reduced steps are often needed

Why reduced steps ?

Newton step with $t \in]0, 1]$:

$$\nabla \mathbf{r}(\mathbf{w}) \Delta \mathbf{w} = -\mathbf{r}(\mathbf{w})$$

$$\mathbf{w} \leftarrow \mathbf{w} + t \Delta \mathbf{w}$$

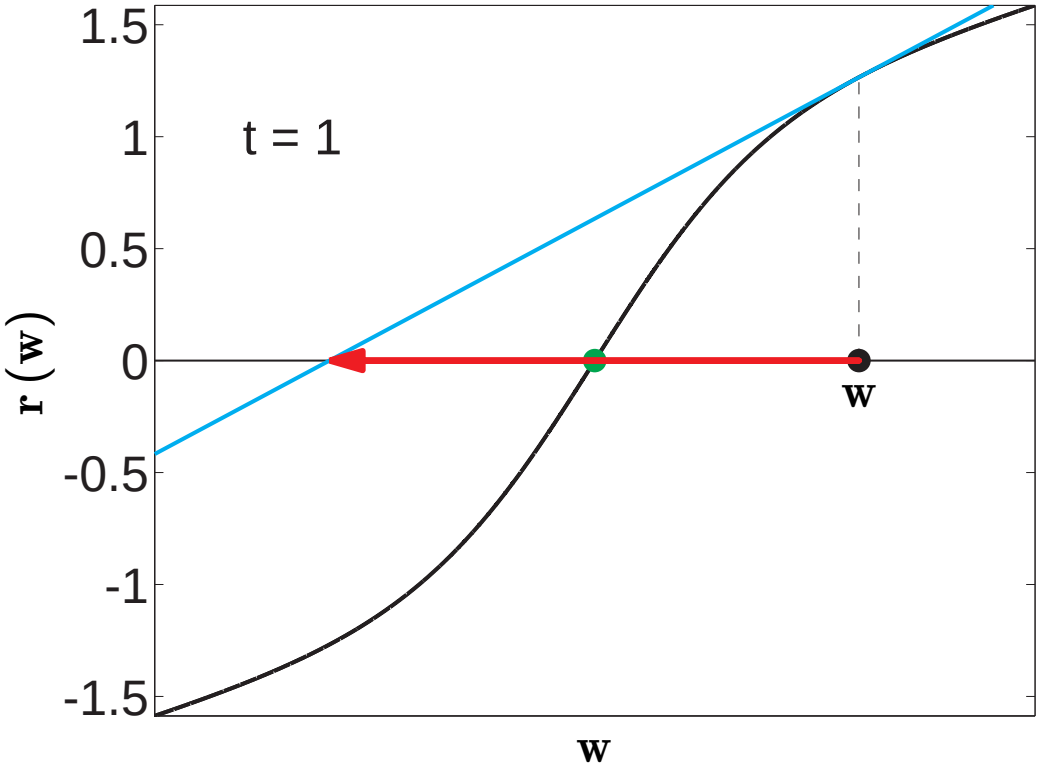


Why reduced steps ?

Newton step with $t \in]0, 1]$:

$$\nabla \mathbf{r}(\mathbf{w}) \Delta \mathbf{w} = -\mathbf{r}(\mathbf{w})$$

$$\mathbf{w} \leftarrow \mathbf{w} + t \Delta \mathbf{w}$$

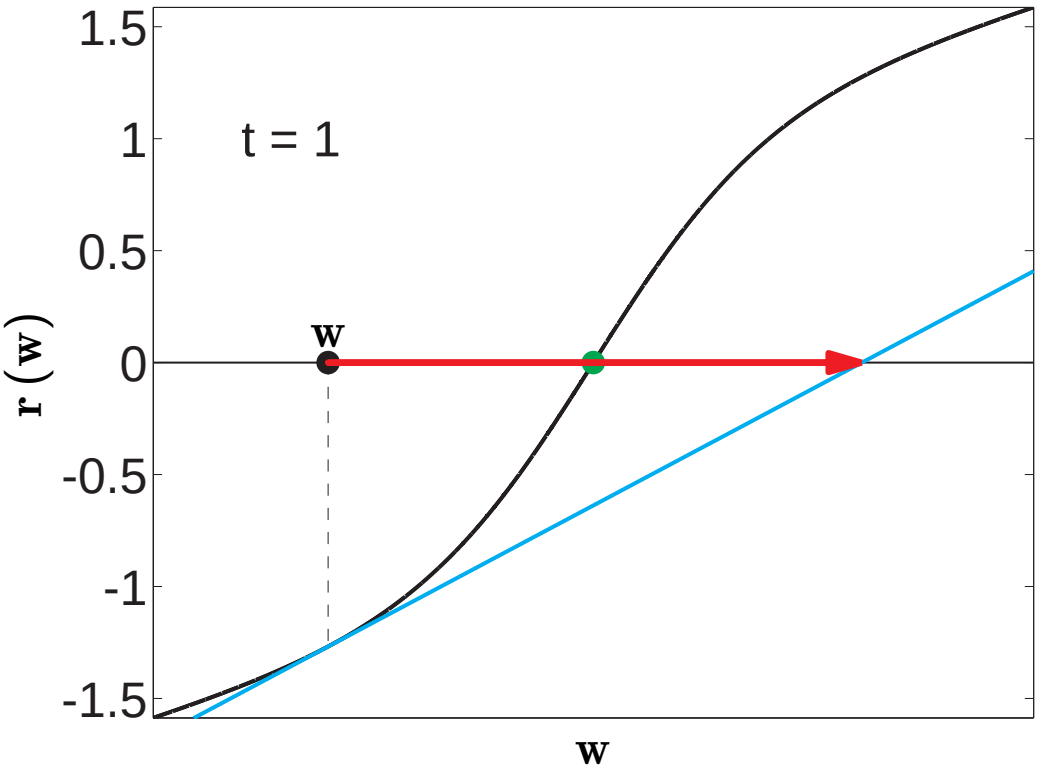


Why reduced steps ?

Newton step with $t \in]0, 1]$:

$$\nabla \mathbf{r}(\mathbf{w}) \Delta \mathbf{w} = -\mathbf{r}(\mathbf{w})$$

$$\mathbf{w} \leftarrow \mathbf{w} + t \Delta \mathbf{w}$$

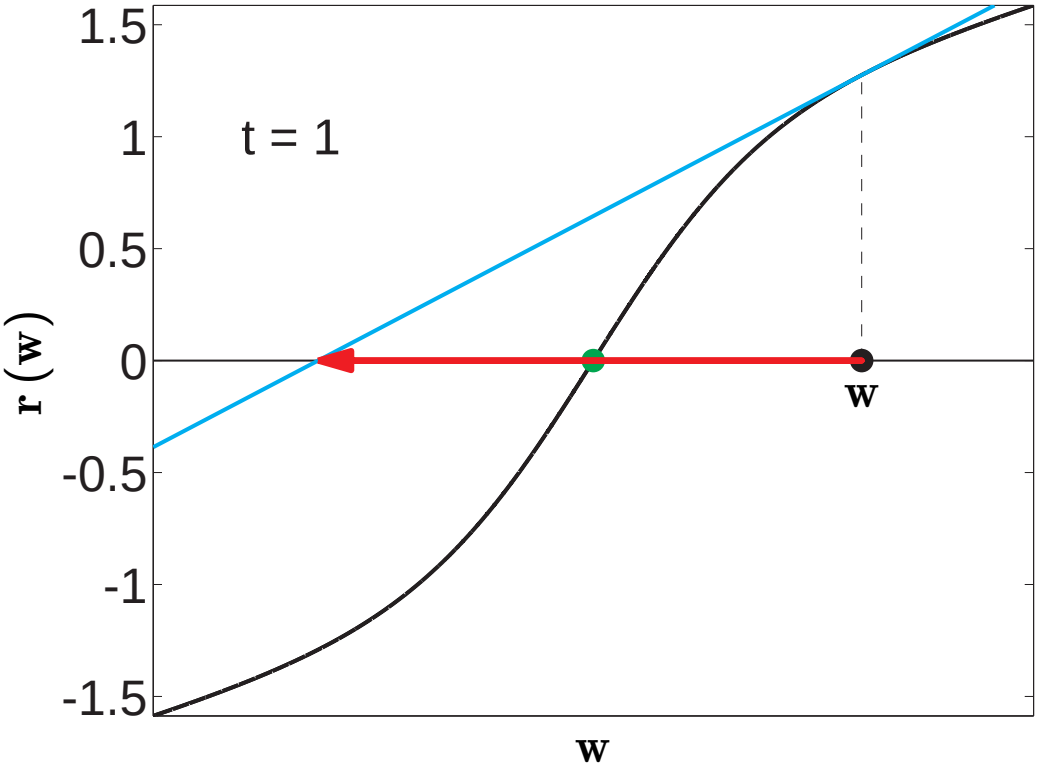


Why reduced steps ?

Newton step with $t \in]0, 1]$:

$$\nabla \mathbf{r}(\mathbf{w}) \Delta \mathbf{w} = -\mathbf{r}(\mathbf{w})$$

$$\mathbf{w} \leftarrow \mathbf{w} + t \Delta \mathbf{w}$$

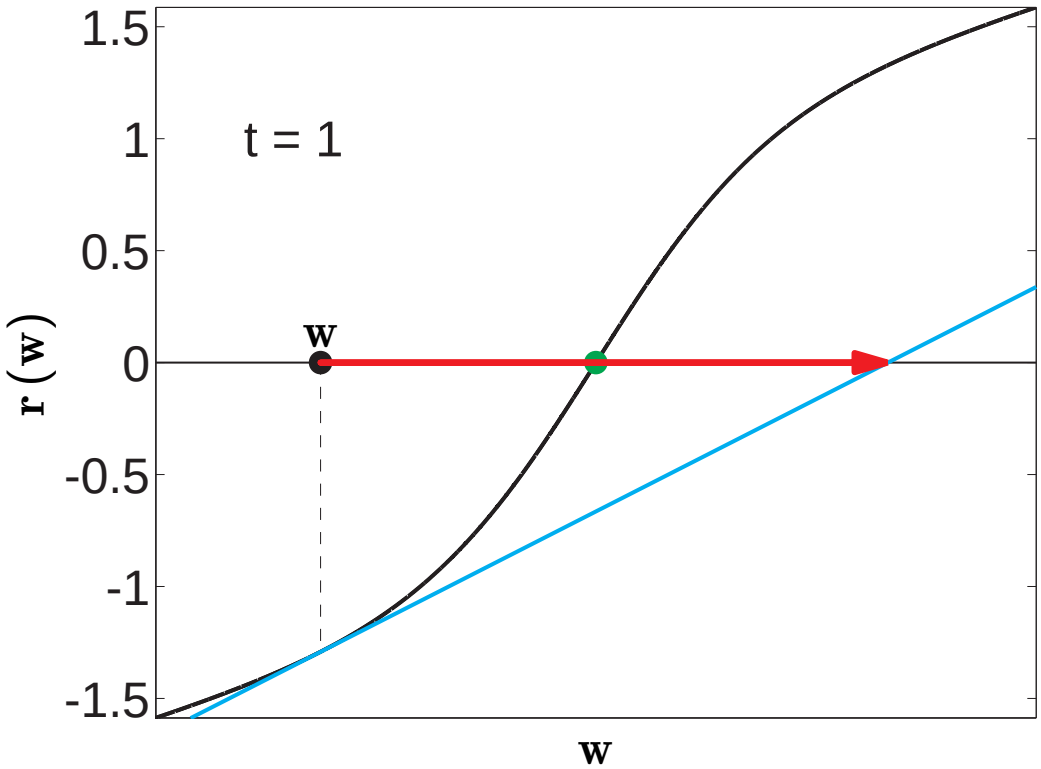


Why reduced steps ?

Newton step with $t \in]0, 1]$:

$$\nabla \mathbf{r}(\mathbf{w}) \Delta \mathbf{w} = -\mathbf{r}(\mathbf{w})$$

$$\mathbf{w} \leftarrow \mathbf{w} + t \Delta \mathbf{w}$$

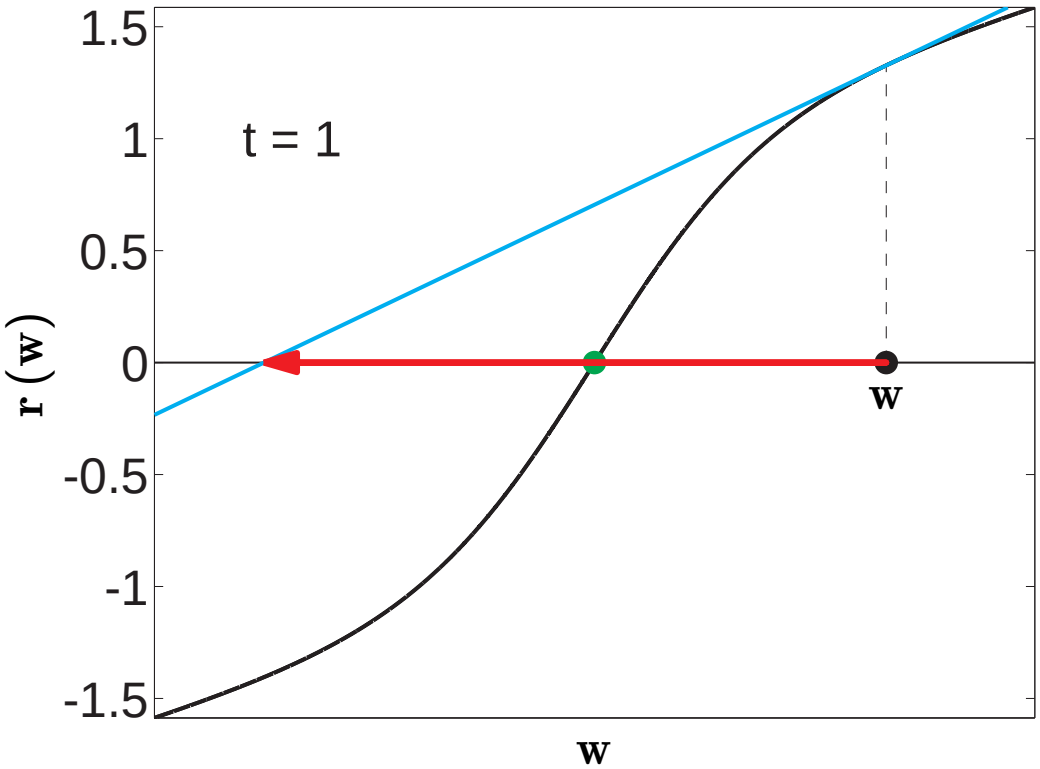


Why reduced steps ?

Newton step with $t \in]0, 1]$:

$$\nabla \mathbf{r}(\mathbf{w}) \Delta \mathbf{w} = -\mathbf{r}(\mathbf{w})$$

$$\mathbf{w} \leftarrow \mathbf{w} + t \Delta \mathbf{w}$$

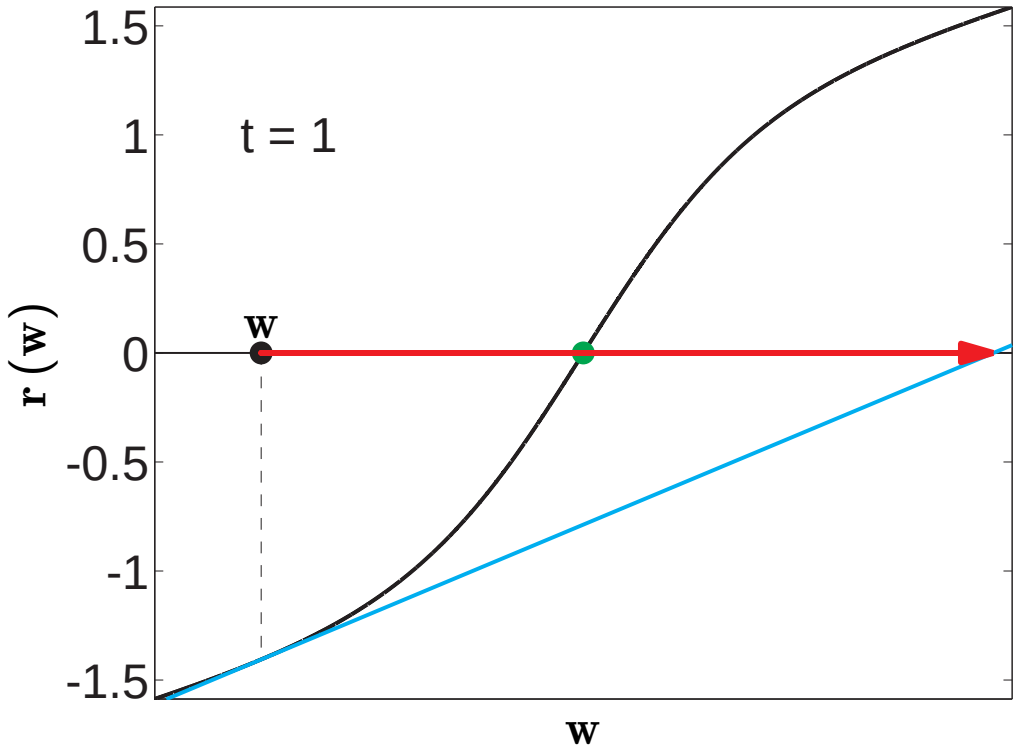


Why reduced steps ?

Newton step with $t \in]0, 1]$:

$$\nabla \mathbf{r}(\mathbf{w}) \Delta \mathbf{w} = -\mathbf{r}(\mathbf{w})$$

$$\mathbf{w} \leftarrow \mathbf{w} + t \Delta \mathbf{w}$$

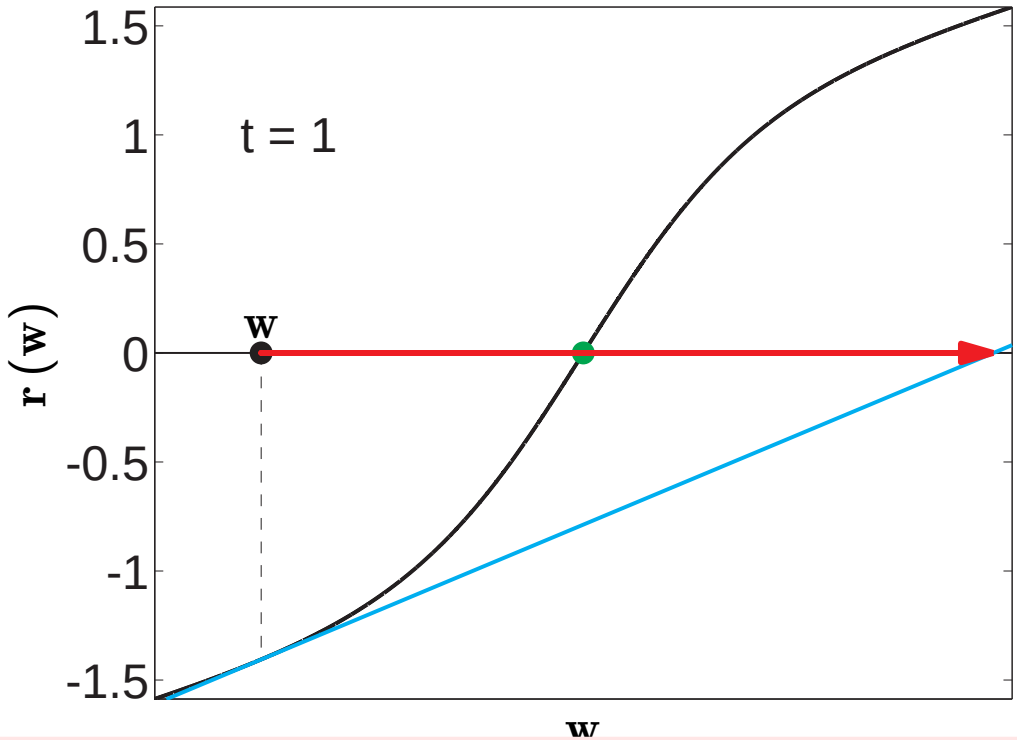


Why reduced steps ?

Newton step with $t \in]0, 1]$:

$$\nabla \mathbf{r}(\mathbf{w}) \Delta \mathbf{w} = -\mathbf{r}(\mathbf{w})$$

$$\mathbf{w} \leftarrow \mathbf{w} + t \Delta \mathbf{w}$$



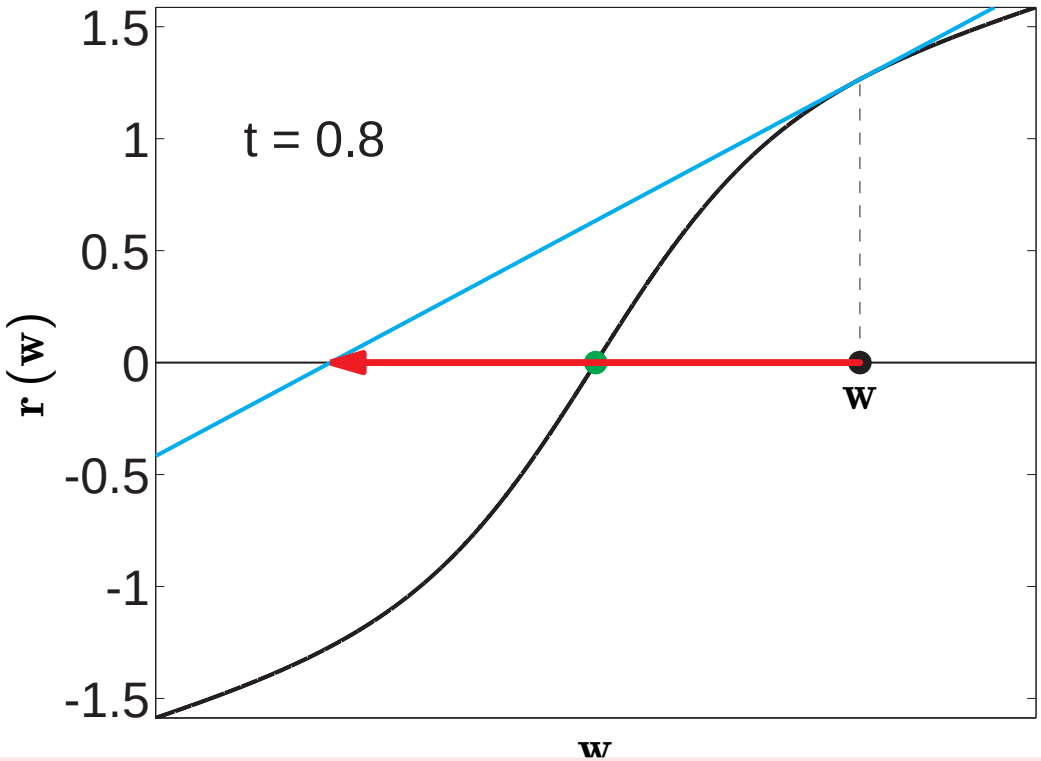
The full-step Newton iteration can be unstable !!

Why reduced steps ?

Newton step with $t \in]0, 1]$:

$$\nabla \mathbf{r}(\mathbf{w}) \Delta \mathbf{w} = -\mathbf{r}(\mathbf{w})$$

$$\mathbf{w} \leftarrow \mathbf{w} + t \Delta \mathbf{w}$$



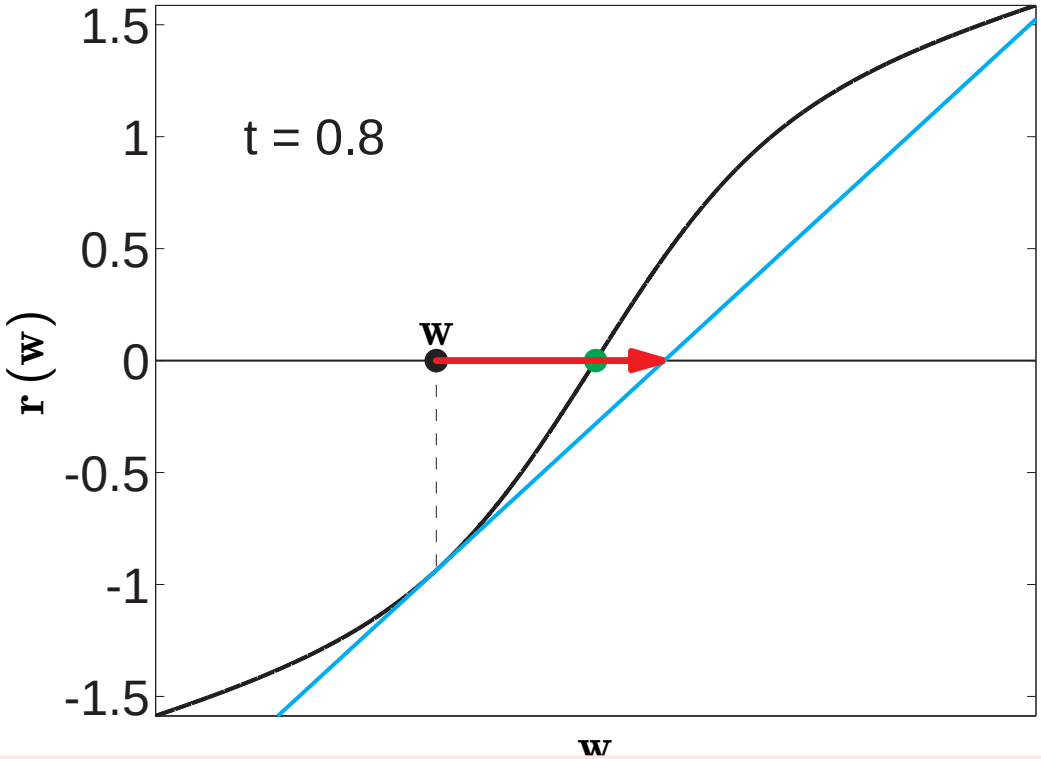
The full-step Newton iteration can be unstable !!

Why reduced steps ?

Newton step with $t \in]0, 1]$:

$$\nabla \mathbf{r}(\mathbf{w}) \Delta \mathbf{w} = -\mathbf{r}(\mathbf{w})$$

$$\mathbf{w} \leftarrow \mathbf{w} + t \Delta \mathbf{w}$$



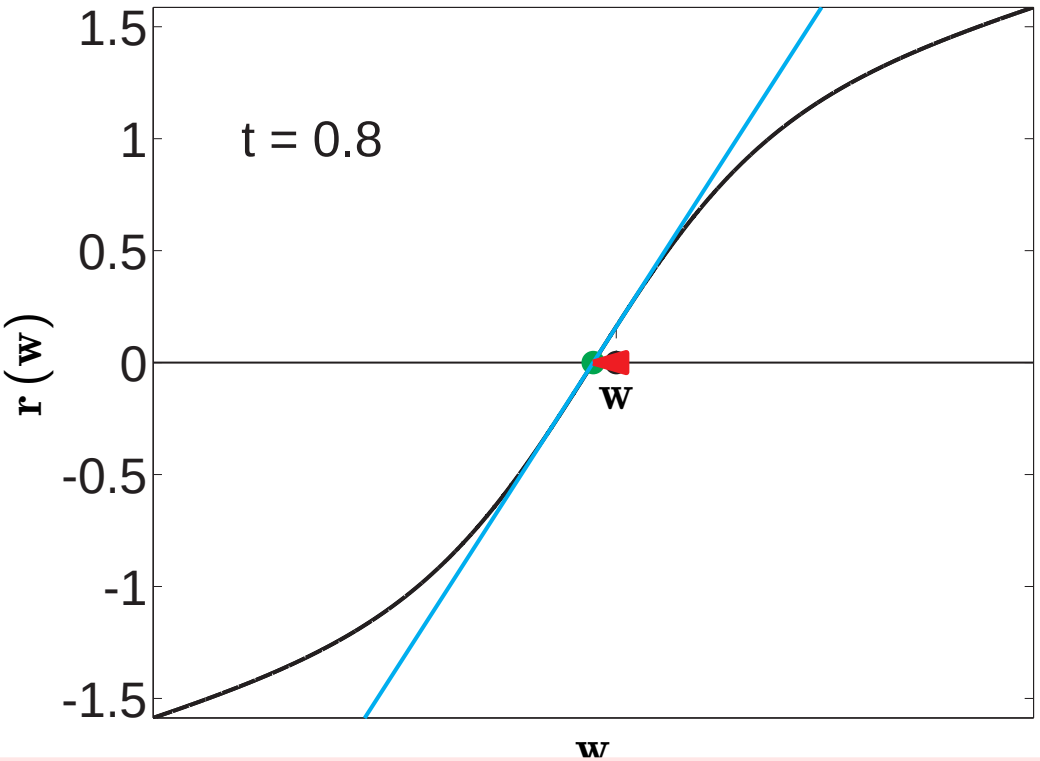
The full-step Newton iteration can be unstable !!

Why reduced steps ?

Newton step with $t \in]0, 1]$:

$$\nabla \mathbf{r}(\mathbf{w}) \Delta \mathbf{w} = -\mathbf{r}(\mathbf{w})$$

$$\mathbf{w} \leftarrow \mathbf{w} + t \Delta \mathbf{w}$$



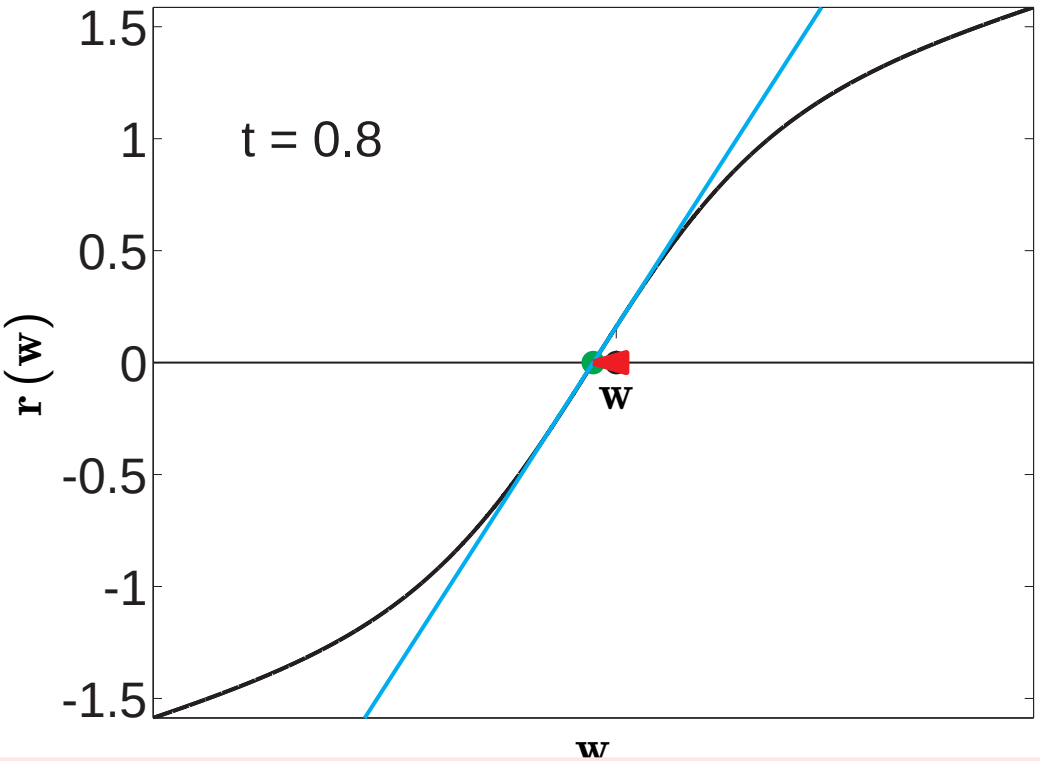
The full-step Newton iteration can be unstable !!

Why reduced steps ?

Newton step with $t \in]0, 1]$:

$$\nabla \mathbf{r}(\mathbf{w}) \Delta \mathbf{w} = -\mathbf{r}(\mathbf{w})$$

$$\mathbf{w} \leftarrow \mathbf{w} + t \Delta \mathbf{w}$$



The full-step Newton iteration can be unstable !! Newton iteration with adequately reduced steps converges (does it ?)

Does Newton always work ?

Is there always a reduced Newton step $t\Delta\mathbf{w}$ "improving" $\mathbf{r}(\mathbf{w})$?

Does Newton always work ?

Is there always a reduced Newton step $t\Delta\mathbf{w}$ "improving" $\mathbf{r}(\mathbf{w})$?
I.e. is there always a $t > 0$ s.t. $\|\mathbf{r}(\mathbf{w} + t\Delta\mathbf{w})\| < \|\mathbf{r}(\mathbf{w})\|$?

Does Newton always work ?

Is there always a reduced Newton step $t\Delta\mathbf{w}$ "improving" $\mathbf{r}(\mathbf{w})$?
I.e. is there always a $t > 0$ s.t. $\|\mathbf{r}(\mathbf{w} + t\Delta\mathbf{w})\| < \|\mathbf{r}(\mathbf{w})\|$? **Yes... sort of**

Does Newton always work ?

Is there always a reduced Newton step $t\Delta\mathbf{w}$ "improving" $\mathbf{r}(\mathbf{w})$?
I.e. is there always a $t > 0$ s.t. $\|\mathbf{r}(\mathbf{w} + t\Delta\mathbf{w})\| < \|\mathbf{r}(\mathbf{w})\|$? **Yes... sort of**

Proof: $\|\mathbf{r}(\mathbf{w} + t\Delta\mathbf{w})\| < \|\mathbf{r}(\mathbf{w})\|$ holds for some $t > 0$ **if**

$$\frac{\mathrm{d}}{\mathrm{d}t}\|\mathbf{r}(\mathbf{w} + t\Delta\mathbf{w})\|_2^2\Big|_{t=0} < 0$$

with $\|\mathbf{r}(\mathbf{w})\|_2^2$ differentiable.

Does Newton always work ?

Is there always a reduced Newton step $t\Delta\mathbf{w}$ "improving" $\mathbf{r}(\mathbf{w})$?
I.e. is there always a $t > 0$ s.t. $\|\mathbf{r}(\mathbf{w} + t\Delta\mathbf{w})\| < \|\mathbf{r}(\mathbf{w})\|$? **Yes... sort of**

Proof: $\|\mathbf{r}(\mathbf{w} + t\Delta\mathbf{w})\| < \|\mathbf{r}(\mathbf{w})\|$ holds for some $t > 0$ **if**

$$\frac{\mathrm{d}}{\mathrm{d}t}\|\mathbf{r}(\mathbf{w} + t\Delta\mathbf{w})\|_2^2\Big|_{t=0} < 0$$

with $\|\mathbf{r}(\mathbf{w})\|_2^2$ differentiable. I.e.

$$2\mathbf{r}(\mathbf{w})^\top \frac{\mathrm{d}}{\mathrm{d}t}\mathbf{r}(\mathbf{w} + t\Delta\mathbf{w})_{t=0} < 0$$

Does Newton always work ?

Is there always a reduced Newton step $t\Delta\mathbf{w}$ "improving" $\mathbf{r}(\mathbf{w})$?
I.e. is there always a $t > 0$ s.t. $\|\mathbf{r}(\mathbf{w} + t\Delta\mathbf{w})\| < \|\mathbf{r}(\mathbf{w})\|$? **Yes... sort of**

Proof: $\|\mathbf{r}(\mathbf{w} + t\Delta\mathbf{w})\| < \|\mathbf{r}(\mathbf{w})\|$ holds for some $t > 0$ **if**

$$\frac{\mathrm{d}}{\mathrm{d}t}\|\mathbf{r}(\mathbf{w} + t\Delta\mathbf{w})\|_2^2\Big|_{t=0} < 0$$

with $\|\mathbf{r}(\mathbf{w})\|_2^2$ differentiable. I.e.

$$2\mathbf{r}(\mathbf{w})^\top \frac{\mathrm{d}}{\mathrm{d}t}\mathbf{r}(\mathbf{w} + t\Delta\mathbf{w})_{t=0} < 0$$

We have

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{r}(\mathbf{w} + t\Delta\mathbf{w})_{t=0} = \nabla\mathbf{r}(\mathbf{w})^\top\Delta\mathbf{w} = -\nabla\mathbf{r}(\mathbf{w})^\top\nabla\mathbf{r}(\mathbf{w})^\top\mathbf{r}(\mathbf{w}) = -\mathbf{r}(\mathbf{w})$$

Does Newton always work ?

Is there always a reduced Newton step $t\Delta\mathbf{w}$ "improving" $\mathbf{r}(\mathbf{w})$?
I.e. is there always a $t > 0$ s.t. $\|\mathbf{r}(\mathbf{w} + t\Delta\mathbf{w})\| < \|\mathbf{r}(\mathbf{w})\|$? **Yes... sort of**

Proof: $\|\mathbf{r}(\mathbf{w} + t\Delta\mathbf{w})\| < \|\mathbf{r}(\mathbf{w})\|$ holds for some $t > 0$ **if**

$$\frac{\mathrm{d}}{\mathrm{d}t}\|\mathbf{r}(\mathbf{w} + t\Delta\mathbf{w})\|_2^2\Big|_{t=0} < 0$$

with $\|\mathbf{r}(\mathbf{w})\|_2^2$ differentiable. I.e.

$$2\mathbf{r}(\mathbf{w})^\top \frac{\mathrm{d}}{\mathrm{d}t}\mathbf{r}(\mathbf{w} + t\Delta\mathbf{w})_{t=0} < 0$$

We have

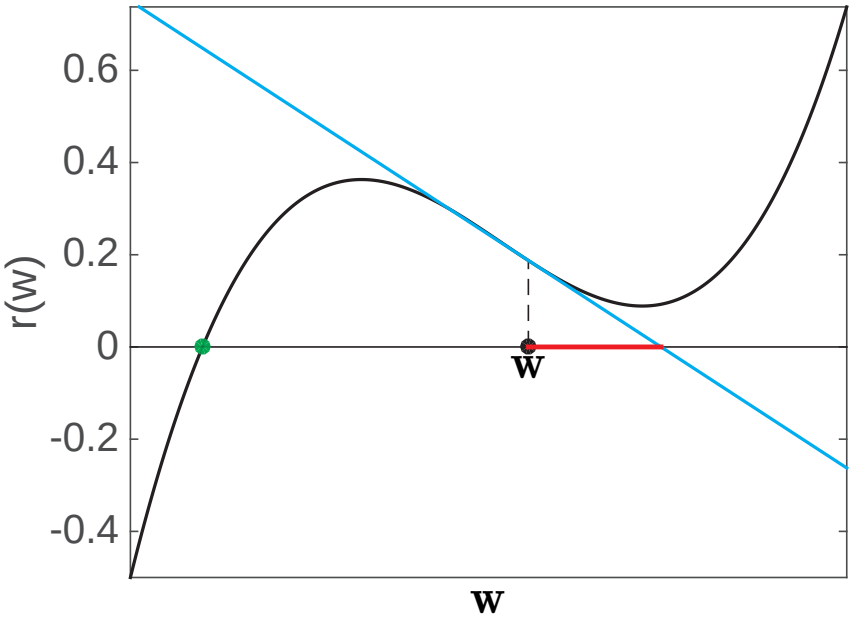
$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{r}(\mathbf{w} + t\Delta\mathbf{w})_{t=0} = \nabla\mathbf{r}(\mathbf{w})^\top\Delta\mathbf{w} = -\nabla\mathbf{r}(\mathbf{w})^\top\nabla\mathbf{r}(\mathbf{w})^\top\mathbf{r}(\mathbf{w}) = -\mathbf{r}(\mathbf{w})$$

Then

$$\frac{\mathrm{d}}{\mathrm{d}t}\|\mathbf{r}(\mathbf{w} + t\Delta\mathbf{w})\|_2^2\Big|_{t=0} = -2\|\mathbf{r}(\mathbf{w})\|_2^2 < 0$$

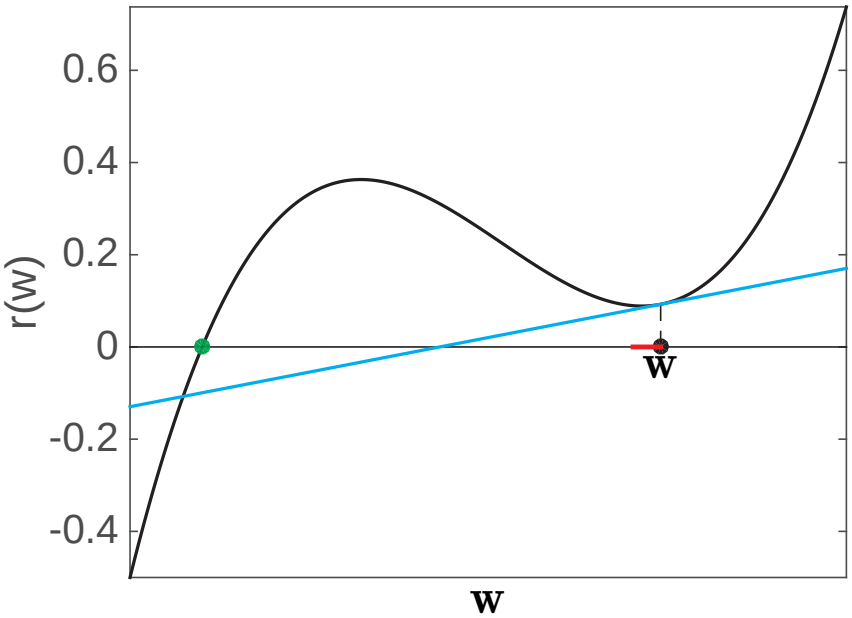
But still, Newton can fail...

Solve $\mathbf{r}(\mathbf{w}) = 0$



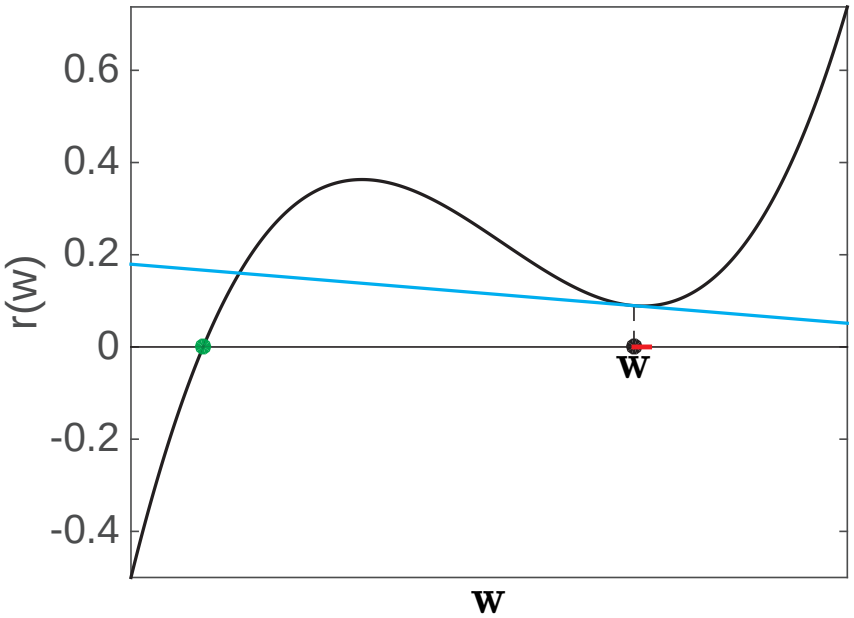
But still, Newton can fail...

Solve $\mathbf{r}(\mathbf{w}) = 0$



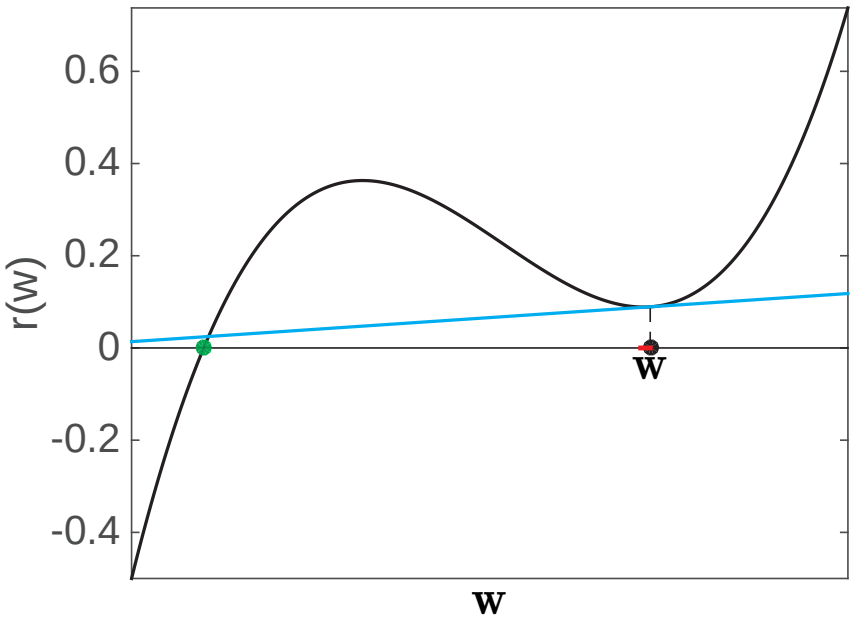
But still, Newton can fail...

Solve $\mathbf{r}(\mathbf{w}) = 0$



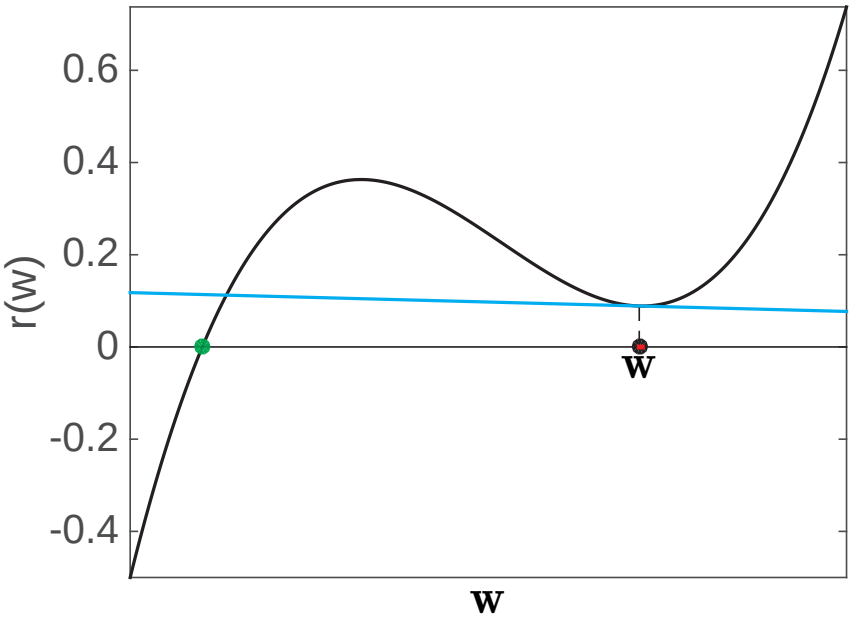
But still, Newton can fail...

Solve $\mathbf{r}(\mathbf{w}) = 0$



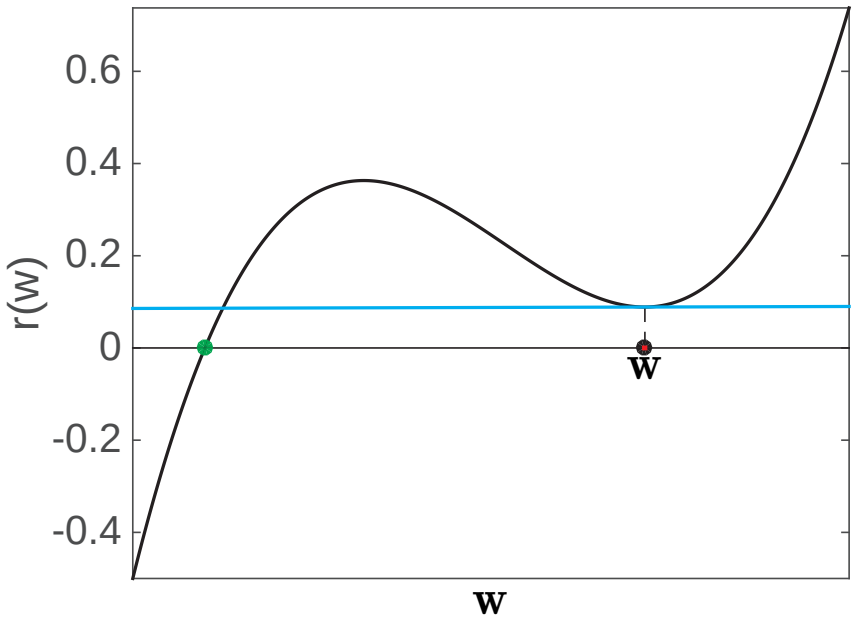
But still, Newton can fail...

Solve $\mathbf{r}(\mathbf{w}) = 0$



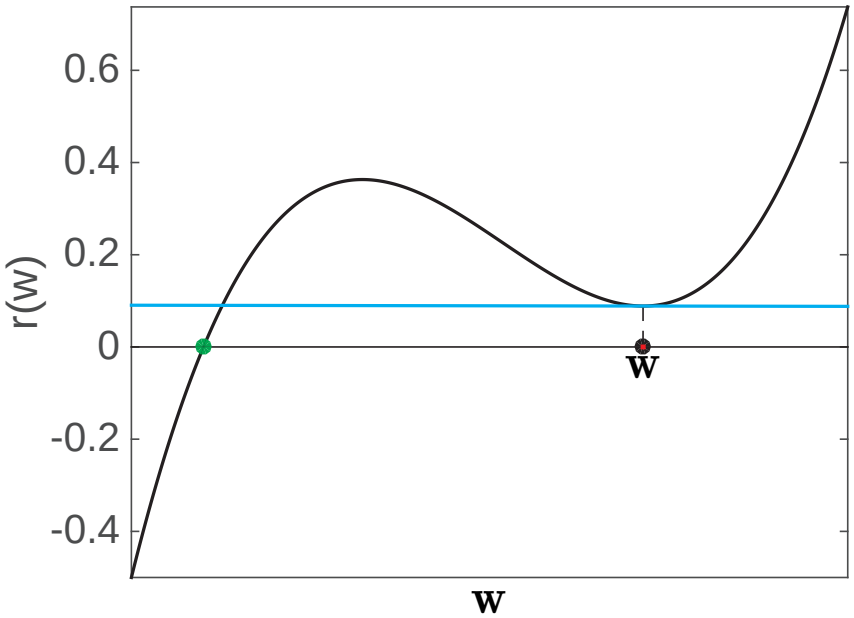
But still, Newton can fail...

Solve $\mathbf{r}(\mathbf{w}) = 0$



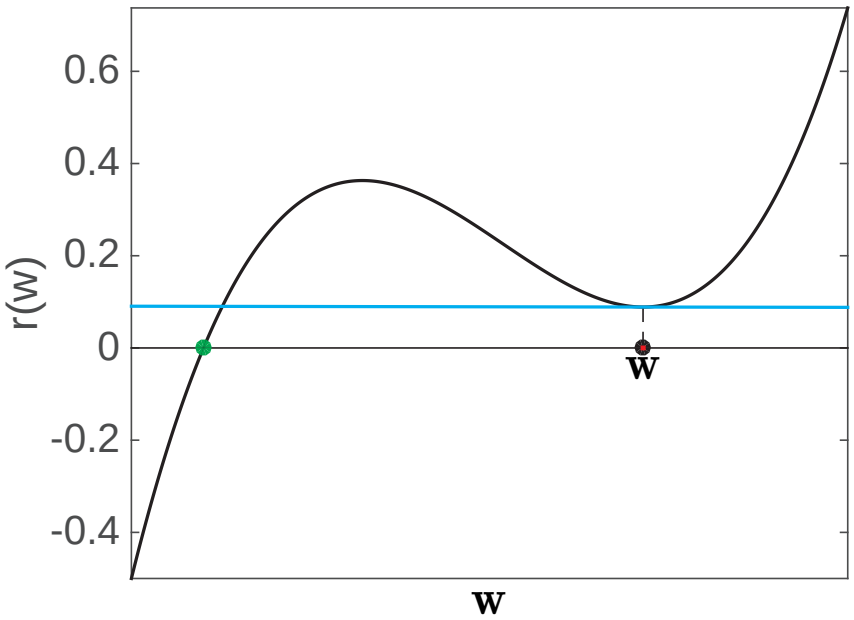
But still, Newton can fail...

Solve $\mathbf{r}(\mathbf{w}) = 0$



But still, Newton can fail...

Solve $\mathbf{r}(\mathbf{w}) = 0$



Newton stops with

$$\mathbf{r}(\mathbf{w}) \neq 0 \text{ and } \nabla \mathbf{r}(\mathbf{w}) \textbf{ singular}$$

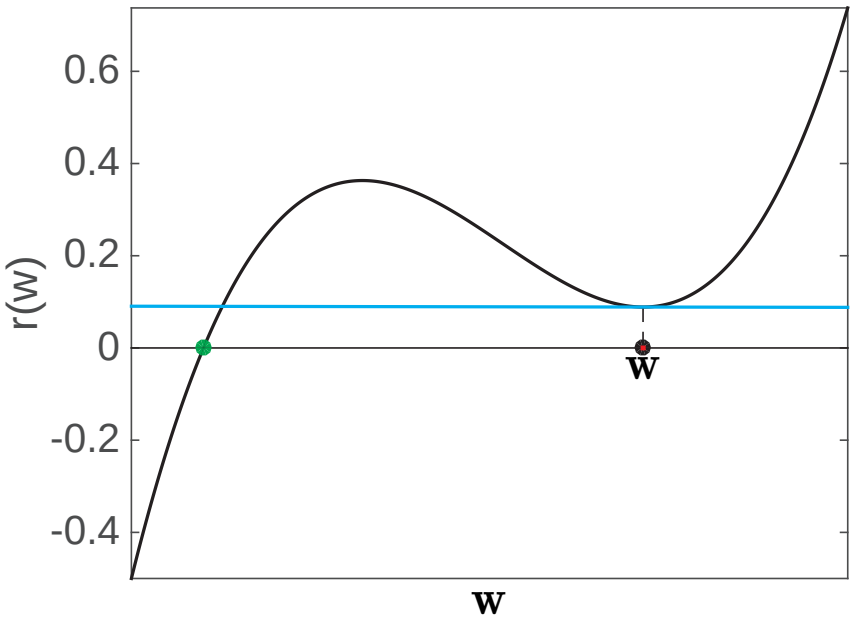
i.e. the Newton direction $\Delta \mathbf{w}$ given by

$$\nabla \mathbf{r}(\mathbf{w}) \Delta \mathbf{w} = -\mathbf{r}(\mathbf{w})$$

is undefined

But still, Newton can fail...

Solve $\mathbf{r}(\mathbf{w}) = 0$



Newton stops with

$$\mathbf{r}(\mathbf{w}) \neq 0 \text{ and } \nabla \mathbf{r}(\mathbf{w}) \textbf{ singular}$$

i.e. the Newton direction $\Delta \mathbf{w}$ given by

$$\nabla \mathbf{r}(\mathbf{w}) \Delta \mathbf{w} = -\mathbf{r}(\mathbf{w})$$

is undefined

Reduced-step Newton iteration converges to a solution of $\mathbf{r}(\mathbf{w}) = 0$ if $\nabla \mathbf{r}$ is full rank everywhere

Convergence of Newton methods & Inexact Jacobian

Consider $\mathbf{r}(\mathbf{w}^*) = 0$, and the Newton iteration $\mathbf{w}^+ = \mathbf{w} - M^{-1}\mathbf{r}(\mathbf{w})$.

Convergence of Newton methods & Inexact Jacobian

Consider $\mathbf{r}(\mathbf{w}^*) = 0$, and the Newton iteration $\mathbf{w}^+ = \mathbf{w} - M^{-1}\mathbf{r}(\mathbf{w})$. We then have:

$$\mathbf{w}^+ - \mathbf{w}^* = \mathbf{w} - \mathbf{w}^* - M^{-1}\left(\mathbf{r}(\mathbf{w}) - \mathbf{r}(\mathbf{w}^*)\right)$$

Convergence of Newton methods & Inexact Jacobian

Consider $\mathbf{r}(\mathbf{w}^*) = 0$, and the Newton iteration $\mathbf{w}^+ = \mathbf{w} - M^{-1}\mathbf{r}(\mathbf{w})$. We then have:

$$\mathbf{w}^+ - \mathbf{w}^* = \mathbf{w} - \mathbf{w}^* - M^{-1}\left(\mathbf{r}(\mathbf{w}) - \mathbf{r}(\mathbf{w}^*)\right)$$

We use the following result from analysis:

$$\mathbf{r}(\mathbf{w}) - \mathbf{r}(\mathbf{w}^*) = \left(\int_0^1 \nabla \mathbf{r}\left(\mathbf{w} + t\left(\mathbf{w}^* - \mathbf{w}\right)\right)^\top \cdot \mathrm{d}t\right)\left(\mathbf{w} - \mathbf{w}^*\right)$$

Convergence of Newton methods & Inexact Jacobian

Consider $\mathbf{r}(\mathbf{w}^*) = 0$, and the Newton iteration $\mathbf{w}^+ = \mathbf{w} - M^{-1}\mathbf{r}(\mathbf{w})$. We then have:

$$\mathbf{w}^+ - \mathbf{w}^* = \mathbf{w} - \mathbf{w}^* - M^{-1}\left(\mathbf{r}(\mathbf{w}) - \mathbf{r}(\mathbf{w}^*)\right)$$

We use the following result from analysis:

$$\mathbf{r}(\mathbf{w}) - \mathbf{r}(\mathbf{w}^*) = \left(\int_0^1 \nabla \mathbf{r}\left(\mathbf{w} + t\left(\mathbf{w}^* - \mathbf{w}\right)\right)^\top \cdot \mathrm{d}t\right)\left(\mathbf{w} - \mathbf{w}^*\right)$$

To obtain: $\mathbf{w}^+ - \mathbf{w}^* = \left(I - M^{-1}\int_0^1 \nabla \mathbf{r}\left(\mathbf{w} + t\left(\mathbf{w}^* - \mathbf{w}\right)\right)^\top \mathrm{d}t\right)\left(\mathbf{w} - \mathbf{w}^*\right)$

Convergence of Newton methods & Inexact Jacobian

Consider $\mathbf{r}(\mathbf{w}^*) = 0$, and the Newton iteration $\mathbf{w}^+ = \mathbf{w} - M^{-1}\mathbf{r}(\mathbf{w})$. We then have:

$$\mathbf{w}^+ - \mathbf{w}^* = \mathbf{w} - \mathbf{w}^* - M^{-1}\left(\mathbf{r}(\mathbf{w}) - \mathbf{r}(\mathbf{w}^*)\right)$$

We use the following result from analysis:

$$\mathbf{r}(\mathbf{w}) - \mathbf{r}(\mathbf{w}^*) = \left(\int_0^1 \nabla \mathbf{r}\left(\mathbf{w} + t\left(\mathbf{w}^* - \mathbf{w}\right)\right)^\top \cdot \mathrm{d}t\right)\left(\mathbf{w} - \mathbf{w}^*\right)$$

To obtain: $\mathbf{w}^+ - \mathbf{w}^* = \left(I - M^{-1}\int_0^1 \nabla \mathbf{r}\left(\mathbf{w} + t\left(\mathbf{w}^* - \mathbf{w}\right)\right)^\top \mathrm{d}t\right)\left(\mathbf{w} - \mathbf{w}^*\right)$

And equivalently:

$$\mathbf{w}^+ - \mathbf{w}^* = M^{-1}\left(M - \nabla \mathbf{r}(\mathbf{w})^\top - \int_0^1 \nabla \mathbf{r}\left(\mathbf{w} + t\left(\mathbf{w}^* - \mathbf{w}\right)\right)^\top - \nabla \mathbf{r}(\mathbf{w})^\top \mathrm{d}t\right)\left(\mathbf{w} - \mathbf{w}^*\right)$$

Convergence of Newton methods & Inexact Jacobian

Consider $\mathbf{r}(\mathbf{w}^*) = 0$, and the Newton iteration $\mathbf{w}^+ = \mathbf{w} - M^{-1}\mathbf{r}(\mathbf{w})$. We then have:

$$\mathbf{w}^+ - \mathbf{w}^* = \mathbf{w} - \mathbf{w}^* - M^{-1}\left(\mathbf{r}(\mathbf{w}) - \mathbf{r}(\mathbf{w}^*)\right)$$

We use the following result from analysis:

$$\mathbf{r}(\mathbf{w}) - \mathbf{r}(\mathbf{w}^*) = \left(\int_0^1 \nabla \mathbf{r}\left(\mathbf{w} + t\left(\mathbf{w}^* - \mathbf{w}\right)\right)^\top \cdot \mathrm{d}t\right)\left(\mathbf{w} - \mathbf{w}^*\right)$$

To obtain: $\mathbf{w}^+ - \mathbf{w}^* = \left(I - M^{-1}\int_0^1 \nabla \mathbf{r}\left(\mathbf{w} + t\left(\mathbf{w}^* - \mathbf{w}\right)\right)^\top \mathrm{d}t\right)\left(\mathbf{w} - \mathbf{w}^*\right)$

And equivalently:

$$\mathbf{w}^+ - \mathbf{w}^* = M^{-1}\left(M - \nabla \mathbf{r}(\mathbf{w})^\top - \int_0^1 \nabla \mathbf{r}\left(\mathbf{w} + t\left(\mathbf{w}^* - \mathbf{w}\right)\right)^\top - \nabla \mathbf{r}(\mathbf{w})^\top \mathrm{d}t\right)\left(\mathbf{w} - \mathbf{w}^*\right)$$

So that:

$$\begin{aligned}\|\mathbf{w}^+ - \mathbf{w}^*\| \leq & \left\|M^{-1}\left(M - \nabla \mathbf{r}(\mathbf{w})^\top\right)\right\| \cdot \|(\mathbf{w} - \mathbf{w}^*)\| \\ & + \left\|M^{-1}\left(\int_0^1 \nabla \mathbf{r}\left(\mathbf{w} + t\left(\mathbf{w}^* - \mathbf{w}\right)\right)^\top - \nabla \mathbf{r}(\mathbf{w})^\top dt\right)\right\| \cdot \|(\mathbf{w} - \mathbf{w}^*)\|\end{aligned}$$

Convergence of Newton methods & Inexact Jacobian

Consider $\mathbf{r}(\mathbf{w}^*) = 0$, and the Newton iteration $\mathbf{w}^+ = \mathbf{w} - M^{-1}\mathbf{r}(\mathbf{w})$. We then have:

$$\mathbf{w}^+ - \mathbf{w}^* = \mathbf{w} - \mathbf{w}^* - M^{-1}\left(\mathbf{r}(\mathbf{w}) - \mathbf{r}(\mathbf{w}^*)\right)$$

We use the following result from analysis:

$$\mathbf{r}(\mathbf{w}) - \mathbf{r}(\mathbf{w}^*) = \left(\int_0^1 \nabla \mathbf{r}\left(\mathbf{w} + t\left(\mathbf{w}^* - \mathbf{w}\right)\right)^\top \cdot \mathrm{d}t\right)\left(\mathbf{w} - \mathbf{w}^*\right)$$

To obtain: $\mathbf{w}^+ - \mathbf{w}^* = \left(I - M^{-1}\int_0^1 \nabla \mathbf{r}\left(\mathbf{w} + t\left(\mathbf{w}^* - \mathbf{w}\right)\right)^\top \mathrm{d}t\right)\left(\mathbf{w} - \mathbf{w}^*\right)$

And equivalently:

$$\mathbf{w}^+ - \mathbf{w}^* = M^{-1}\left(M - \nabla \mathbf{r}(\mathbf{w})^\top - \int_0^1 \nabla \mathbf{r}\left(\mathbf{w} + t\left(\mathbf{w}^* - \mathbf{w}\right)\right)^\top - \nabla \mathbf{r}(\mathbf{w})^\top \mathrm{d}t\right)\left(\mathbf{w} - \mathbf{w}^*\right)$$

So that:

$$\|\mathbf{w}^+ - \mathbf{w}^*\| \leq \underbrace{\left\|M^{-1}\left(M - \nabla \mathbf{r}(\mathbf{w})^\top\right)\right\|}_{\text{small if } M \approx \nabla \mathbf{r}(\mathbf{w})^\top} \cdot \|(\mathbf{w} - \mathbf{w}^*)\|$$

$$+ \left\|M^{-1}\left(\int_0^1 \nabla \mathbf{r}\left(\mathbf{w} + t\left(\mathbf{w}^* - \mathbf{w}\right)\right)^\top - \nabla \mathbf{r}(\mathbf{w})^\top dt\right)\right\| \cdot \|(\mathbf{w} - \mathbf{w}^*)\|$$

Convergence of Newton methods & Inexact Jacobian

Consider $\mathbf{r}(\mathbf{w}^*) = 0$, and the Newton iteration $\mathbf{w}^+ = \mathbf{w} - M^{-1}\mathbf{r}(\mathbf{w})$. We then have:

$$\mathbf{w}^+ - \mathbf{w}^* = \mathbf{w} - \mathbf{w}^* - M^{-1}\left(\mathbf{r}(\mathbf{w}) - \mathbf{r}(\mathbf{w}^*)\right)$$

We use the following result from analysis:

$$\mathbf{r}(\mathbf{w}) - \mathbf{r}(\mathbf{w}^*) = \left(\int_0^1 \nabla \mathbf{r}\left(\mathbf{w} + t\left(\mathbf{w}^* - \mathbf{w}\right)\right)^\top \cdot \mathrm{d}t\right)\left(\mathbf{w} - \mathbf{w}^*\right)$$

To obtain: $\mathbf{w}^+ - \mathbf{w}^* = \left(I - M^{-1}\int_0^1 \nabla \mathbf{r}\left(\mathbf{w} + t\left(\mathbf{w}^* - \mathbf{w}\right)\right)^\top \mathrm{d}t\right)\left(\mathbf{w} - \mathbf{w}^*\right)$

And equivalently:

$$\mathbf{w}^+ - \mathbf{w}^* = M^{-1}\left(M - \nabla \mathbf{r}(\mathbf{w})^\top - \int_0^1 \nabla \mathbf{r}\left(\mathbf{w} + t\left(\mathbf{w}^* - \mathbf{w}\right)\right)^\top - \nabla \mathbf{r}(\mathbf{w})^\top \mathrm{d}t\right)\left(\mathbf{w} - \mathbf{w}^*\right)$$

So that:

$$\|\mathbf{w}^+ - \mathbf{w}^*\| \leq \underbrace{\left\|M^{-1}\left(M - \nabla \mathbf{r}(\mathbf{w})^\top\right)\right\|}_{\text{small if } M - \nabla \mathbf{r}(\mathbf{w})^\top} \cdot \|\mathbf{w} - \mathbf{w}^*\|$$

$$+ \underbrace{\left\|M^{-1}\left(\int_0^1 \nabla \mathbf{r}\left(\mathbf{w} + t\left(\mathbf{w}^* - \mathbf{w}\right)\right)^\top - \nabla \mathbf{r}(\mathbf{w})^\top dt\right)\right\|}_{\text{small if } \nabla \mathbf{r} \text{ is not changing much between } \mathbf{w} \text{ and } \mathbf{w}^*} \cdot \|\mathbf{w} - \mathbf{w}^*\|$$

Convergence of Newton methods

Theorem: consider the Newton iteration $\mathbf{w}_{k+1} = \mathbf{w}_k - M_k^{-1}\mathbf{r}(\mathbf{w}_k)$ and assume

- Lipschitz condition: $\|M_k^{-1}(\nabla \mathbf{r}(\mathbf{w})^\top - \nabla \mathbf{r}(\mathbf{w}^*))^\top\| \leq \omega \|\mathbf{w} - \mathbf{w}^*\|$

Convergence of Newton methods

Theorem: consider the Newton iteration $\mathbf{w}_{k+1} = \mathbf{w}_k - M_k^{-1}\mathbf{r}(\mathbf{w}_k)$ and assume

- Lipschitz condition: $\|M_k^{-1}(\nabla \mathbf{r}(\mathbf{w})^\top - \nabla \mathbf{r}(\mathbf{w}^*))^\top\| \leq \omega \|\mathbf{w} - \mathbf{w}^*\|$
- Bound on the Jacobian approximation error $\|M_k^{-1}(\nabla \mathbf{r}(\mathbf{w}_k)^\top - M_k)\| \leq \kappa_k < \kappa < 1$

Convergence of Newton methods

Theorem: consider the Newton iteration $\mathbf{w}_{k+1} = \mathbf{w}_k - M_k^{-1}\mathbf{r}(\mathbf{w}_k)$ and assume

- Lipschitz condition: $\|M_k^{-1}(\nabla \mathbf{r}(\mathbf{w})^\top - \nabla \mathbf{r}(\mathbf{w}^*))^\top\| \leq \omega \|\mathbf{w} - \mathbf{w}^*\|$
- Bound on the Jacobian approximation error $\|M_k^{-1}(\nabla \mathbf{r}(\mathbf{w}_k)^\top - M_k)\| \leq \kappa_k < \kappa < 1$
- Good initial guess $\|\mathbf{w}_0 - \mathbf{w}^*\| \leq \frac{2}{\omega}(1-\kappa)$

Then $\mathbf{w}_k \rightarrow \mathbf{w}^*$ with the following linear-quadratic contraction rate:

$$\|\mathbf{w}_{k+1} - \mathbf{w}^*\| \quad \leq \quad \left(\kappa_k + \frac{\omega}{2}\|\mathbf{w}_k - \mathbf{w}^*\|\right)\|\mathbf{w}_k - \mathbf{w}^*\|.$$

Convergence of Newton methods

Theorem: consider the Newton iteration $\mathbf{w}_{k+1} = \mathbf{w}_k - M_k^{-1}\mathbf{r}(\mathbf{w}_k)$ and assume

- Lipschitz condition: $\|M_k^{-1}(\nabla \mathbf{r}(\mathbf{w})^\top - \nabla \mathbf{r}(\mathbf{w}^*))^\top\| \leq \omega \|\mathbf{w} - \mathbf{w}^*\|$
- Bound on the Jacobian approximation error $\|M_k^{-1}(\nabla \mathbf{r}(\mathbf{w}_k)^\top - M_k)\| \leq \kappa_k < \kappa < 1$
- Good initial guess $\|\mathbf{w}_0 - \mathbf{w}^*\| \leq \frac{2}{\omega}(1-\kappa)$

Then $\mathbf{w}_k \to \mathbf{w}^*$ with the following linear-quadratic contraction rate:

$$\|\mathbf{w}_{k+1} - \mathbf{w}^*\| \quad \leq \quad \left(\kappa_k + \frac{\omega}{2}\|\mathbf{w}_k - \mathbf{w}^*\|\right)\|\mathbf{w}_k - \mathbf{w}^*\|.$$

Proof

$$\|\mathbf{w}_{k+1} - \mathbf{w}^*\| \leq \overbrace{\left\| M_k^{-1}\left(M_k - \nabla \mathbf{r}(\mathbf{w}_k)^\top\right)\right\|}^{\leq \kappa_k} \cdot \|\mathbf{w}_k - \mathbf{w}^*\|$$
$$+ \left\| M_k^{-1}\left(\int_0^1 \nabla \mathbf{r}(\mathbf{w}_k + t(\mathbf{w}^* - \mathbf{w}_k))^\top - \nabla \mathbf{r}(\mathbf{w}_k)^\top \mathrm{d}t\right)\right\| \cdot \|\mathbf{w}_k - \mathbf{w}^*\|$$

Convergence of Newton methods

Theorem: consider the Newton iteration $\mathbf{w}_{k+1} = \mathbf{w}_k - M_k^{-1}\mathbf{r}(\mathbf{w}_k)$ and assume

- Lipschitz condition: $\|M_k^{-1}(\nabla \mathbf{r}(\mathbf{w})^\top - \nabla \mathbf{r}(\mathbf{w}^*))^\top\| \leq \omega \|\mathbf{w} - \mathbf{w}^*\|$
- Bound on the Jacobian approximation error $\|M_k^{-1}(\nabla \mathbf{r}(\mathbf{w}_k)^\top - M_k)\| \leq \kappa_k < \kappa < 1$
- Good initial guess $\|\mathbf{w}_0 - \mathbf{w}^*\| \leq \frac{2}{\omega}(1-\kappa)$

Then $\mathbf{w}_k \to \mathbf{w}^*$ with the following linear-quadratic contraction rate:

$$\|\mathbf{w}_{k+1} - \mathbf{w}^*\| \quad \leq \quad \left(\kappa_k + \frac{\omega}{2}\|\mathbf{w}_k - \mathbf{w}^*\|\right)\|\mathbf{w}_k - \mathbf{w}^*\|.$$

Proof

$$\|\mathbf{w}_{k+1} - \mathbf{w}^*\| \leq \overbrace{\left\| M_k^{-1}\left(M_k - \nabla \mathbf{r}(\mathbf{w}_k)^\top\right)\right\|}^{\leq \kappa_k} \cdot \|\mathbf{w}_k - \mathbf{w}^*\|$$

$$+ \left\| M_k^{-1}\left(\int_0^1 \nabla \mathbf{r}(\mathbf{w}_k + t(\mathbf{w}^* - \mathbf{w}_k))^\top - \nabla \mathbf{r}(\mathbf{w}_k)^\top \mathrm{d}t\right)\right\| \cdot \|\mathbf{w}_k - \mathbf{w}^*\|$$

and $$\left\| M_k^{-1}\left(\int_0^1 \nabla \mathbf{r}(\mathbf{w}_k + t(\mathbf{w}^* - \mathbf{w}_k))^\top - \nabla \mathbf{r}(\mathbf{w}_k)^\top \mathrm{d}t\right)\right\| \leq$$

$$\int_0^1 \left\| M_k^{-1}\left(\nabla \mathbf{r}(\mathbf{w}_k + t(\mathbf{w}^* - \mathbf{w}_k))^\top - \nabla \mathbf{r}(\mathbf{w}_k)^\top\right)\right\| \mathrm{d}t$$

Convergence of Newton methods

Theorem: consider the Newton iteration $\mathbf{w}_{k+1} = \mathbf{w}_k - M_k^{-1}\mathbf{r}(\mathbf{w}_k)$ and assume

- Lipschitz condition: $\|M_k^{-1}(\nabla \mathbf{r}(\mathbf{w})^\top - \nabla \mathbf{r}(\mathbf{w}^*))^\top\| \leq \omega \|\mathbf{w} - \mathbf{w}^*\|$
- Bound on the Jacobian approximation error $\|M_k^{-1}(\nabla \mathbf{r}(\mathbf{w}_k)^\top - M_k)\| \leq \kappa_k < \kappa < 1$
- Good initial guess $\|\mathbf{w}_0 - \mathbf{w}^*\| \leq \frac{2}{\omega}(1-\kappa)$

Then $\mathbf{w}_k \to \mathbf{w}^*$ with the following linear-quadratic contraction rate:

$$\|\mathbf{w}_{k+1} - \mathbf{w}^*\| \quad \leq \quad \left(\kappa_k + \frac{\omega}{2}\|\mathbf{w}_k - \mathbf{w}^*\|\right)\|\mathbf{w}_k - \mathbf{w}^*\|.$$

Proof

$$\|\mathbf{w}_{k+1} - \mathbf{w}^*\| \leq \overbrace{\left\| M_k^{-1}\left(M_k - \nabla \mathbf{r}(\mathbf{w}_k)^\top\right)\right\|}^{\leq \kappa_k} \cdot \|\mathbf{w}_k - \mathbf{w}^*\|$$

$$+ \left\| M_k^{-1}\left(\int_0^1 \nabla \mathbf{r}(\mathbf{w}_k + t(\mathbf{w}^* - \mathbf{w}_k))^\top - \nabla \mathbf{r}(\mathbf{w}_k)^\top \mathrm{d}t\right)\right\| \cdot \|\mathbf{w}_k - \mathbf{w}^*\|$$

and $$\left\| M_k^{-1}\left(\int_0^1 \nabla \mathbf{r}(\mathbf{w}_k + t(\mathbf{w}^* - \mathbf{w}_k))^\top - \nabla \mathbf{r}(\mathbf{w}_k)^\top \mathrm{d}t\right)\right\| \leq$$

$$\int_0^1 \left\| M_k^{-1}\left(\nabla \mathbf{r}(\mathbf{w}_k + t(\mathbf{w}^* - \mathbf{w}_k))^\top - \nabla \mathbf{r}(\mathbf{w}_k)^\top\right)\right\| \mathrm{d}t$$

$$\leq \int_0^1 \omega \|\mathbf{w}_k + t(\mathbf{w}^* - \mathbf{w}_k) - \mathbf{w}_k\| \mathrm{d}t$$

Convergence of Newton methods

Theorem: consider the Newton iteration $\mathbf{w}_{k+1} = \mathbf{w}_k - M_k^{-1}\mathbf{r}(\mathbf{w}_k)$ and assume

- Lipschitz condition: $\|M_k^{-1}(\nabla \mathbf{r}(\mathbf{w})^\top - \nabla \mathbf{r}(\mathbf{w}^*))^\top\| \leq \omega \|\mathbf{w} - \mathbf{w}^*\|$
- Bound on the Jacobian approximation error $\|M_k^{-1}(\nabla \mathbf{r}(\mathbf{w}_k)^\top - M_k)\| \leq \kappa_k < \kappa < 1$
- Good initial guess $\|\mathbf{w}_0 - \mathbf{w}^*\| \leq \frac{2}{\omega}(1-\kappa)$

Then $\mathbf{w}_k \to \mathbf{w}^*$ with the following linear-quadratic contraction rate:

$$\|\mathbf{w}_{k+1} - \mathbf{w}^*\| \quad \leq \quad \left(\kappa_k + \frac{\omega}{2}\|\mathbf{w}_k - \mathbf{w}^*\|\right)\|\mathbf{w}_k - \mathbf{w}^*\|.$$

Proof

$$\|\mathbf{w}_{k+1} - \mathbf{w}^*\| \leq \overbrace{\left\| M_k^{-1}\left(M_k - \nabla \mathbf{r}(\mathbf{w}_k)^\top\right)\right\|}^{\leq \kappa_k} \cdot \|\mathbf{w}_k - \mathbf{w}^*\|$$

$$+ \left\| M_k^{-1}\left(\int_0^1 \nabla \mathbf{r}(\mathbf{w}_k + t(\mathbf{w}^* - \mathbf{w}_k))^\top - \nabla \mathbf{r}(\mathbf{w}_k)^\top \mathrm{d}t\right)\right\| \cdot \|\mathbf{w}_k - \mathbf{w}^*\|$$

and $$\left\| M_k^{-1}\left(\int_0^1 \nabla \mathbf{r}(\mathbf{w}_k + t(\mathbf{w}^* - \mathbf{w}_k))^\top - \nabla \mathbf{r}(\mathbf{w}_k)^\top \mathrm{d}t\right)\right\| \leq$$

$$\int_0^1 \left\| M_k^{-1}\left(\nabla \mathbf{r}(\mathbf{w}_k + t(\mathbf{w}^* - \mathbf{w}_k))^\top - \nabla \mathbf{r}(\mathbf{w}_k)^\top\right)\right\| \mathrm{d}t$$

$$\leq \int_0^1 \omega \|\mathbf{w}_k + t(\mathbf{w}^* - \mathbf{w}_k) - \mathbf{w}_k\| \mathrm{d}t \leq \int_0^1 t\omega \|\mathbf{w}^* - \mathbf{w}_k\| \mathrm{d}t$$

Convergence of Newton methods

Theorem: consider the Newton iteration $\mathbf{w}_{k+1} = \mathbf{w}_k - M_k^{-1}\mathbf{r}(\mathbf{w}_k)$ and assume

- Lipschitz condition: $\|M_k^{-1}(\nabla \mathbf{r}(\mathbf{w})^\top - \nabla \mathbf{r}(\mathbf{w}^*))^\top\| \leq \omega \|\mathbf{w} - \mathbf{w}^*\|$
- Bound on the Jacobian approximation error $\|M_k^{-1}(\nabla \mathbf{r}(\mathbf{w}_k)^\top - M_k)\| \leq \kappa_k < \kappa < 1$
- Good initial guess $\|\mathbf{w}_0 - \mathbf{w}^*\| \leq \frac{2}{\omega}(1-\kappa)$

Then $\mathbf{w}_k \to \mathbf{w}^*$ with the following linear-quadratic contraction rate:

$$\|\mathbf{w}_{k+1} - \mathbf{w}^*\| \quad \leq \quad \left(\kappa_k + \frac{\omega}{2}\|\mathbf{w}_k - \mathbf{w}^*\|\right)\|\mathbf{w}_k - \mathbf{w}^*\|.$$

Proof

$$\|\mathbf{w}_{k+1} - \mathbf{w}^*\| \leq \overbrace{\left\| M_k^{-1}\left(M_k - \nabla \mathbf{r}(\mathbf{w}_k)^\top\right)\right\|}^{\leq \kappa_k} \cdot \|\mathbf{w}_k - \mathbf{w}^*\|$$

$$+ \left\| M_k^{-1}\left(\int_0^1 \nabla \mathbf{r}(\mathbf{w}_k + t(\mathbf{w}^* - \mathbf{w}_k))^\top - \nabla \mathbf{r}(\mathbf{w}_k)^\top \mathrm{d}t\right)\right\| \cdot \|\mathbf{w}_k - \mathbf{w}^*\|$$

and $$\left\| M_k^{-1}\left(\int_0^1 \nabla \mathbf{r}(\mathbf{w}_k + t(\mathbf{w}^* - \mathbf{w}_k))^\top - \nabla \mathbf{r}(\mathbf{w}_k)^\top \mathrm{d}t\right)\right\| \leq$$

$$\int_0^1 \left\| M_k^{-1}\left(\nabla \mathbf{r}(\mathbf{w}_k + t(\mathbf{w}^* - \mathbf{w}_k))^\top - \nabla \mathbf{r}(\mathbf{w}_k)^\top\right)\right\| \mathrm{d}t$$

$$\leq \int_0^1 \omega \|\mathbf{w}_k + t(\mathbf{w}^* - \mathbf{w}_k) - \mathbf{w}_k\| \mathrm{d}t \leq \int_0^1 t\omega \|\mathbf{w}^* - \mathbf{w}_k\| \mathrm{d}t = \frac{\omega}{2}\|\mathbf{w}^* - \mathbf{w}_k\|$$

Affine invariance of the (exact) Newton method

Affine change of coordinates

Consider: $\mathbf{w} = A\mathbf{v} + \mathbf{a}$ with $A \in \mathbb{R}^{n \times n}$ non-singular and $\mathbf{a} \in \mathbb{R}^n$.

Affine invariance of the (exact) Newton method

Affine change of coordinates

Consider: $\mathbf{w} = A\mathbf{v} + \mathbf{a}$ with $A \in \mathbb{R}^{n \times n}$ non-singular and $\mathbf{a} \in \mathbb{R}^n$.

Define $\tilde{\mathbf{r}}(\mathbf{v}) = \mathbf{r}(A\mathbf{v} + \mathbf{a})$ then: $\nabla_{\mathbf{v}} \tilde{\mathbf{r}}(\mathbf{v}) = A^\top \nabla_{\mathbf{w}} \mathbf{r}(\mathbf{w})$

Affine invariance of the (exact) Newton method

Affine change of coordinates

Consider: $\mathbf{w} = A\mathbf{v} + \mathbf{a}$ with $A \in \mathbb{R}^{n \times n}$ non-singular and $\mathbf{a} \in \mathbb{R}^n$.

Define $\tilde{\mathbf{r}}(\mathbf{v}) = \mathbf{r}(A\mathbf{v} + \mathbf{a})$ then: $$\nabla_{\mathbf{v}} \tilde{\mathbf{r}}(\mathbf{v}) = A^\top \nabla_{\mathbf{w}} \mathbf{r}(\mathbf{w})$$

The Newton step in $\mathbf{v}$ given by:

$$\nabla_{\mathbf{v}} \tilde{\mathbf{r}}(\mathbf{v})^\top \Delta \mathbf{v} = -\tilde{\mathbf{r}}(\mathbf{v})$$

Affine invariance of the (exact) Newton method

Affine change of coordinates

Consider: $\mathbf{w} = A\mathbf{v} + \mathbf{a}$ with $A \in \mathbb{R}^{n \times n}$ non-singular and $\mathbf{a} \in \mathbb{R}^n$.

Define $\tilde{\mathbf{r}}(\mathbf{v}) = \mathbf{r}(A\mathbf{v} + \mathbf{a})$ then: $$\nabla_{\mathbf{v}} \tilde{\mathbf{r}}(\mathbf{v}) = A^\top \nabla_{\mathbf{w}} \mathbf{r}(\mathbf{w})$$

The Newton step in $\mathbf{v}$ given by:

$$\nabla_{\mathbf{v}} \tilde{\mathbf{r}}(\mathbf{v})^\top \Delta\mathbf{v} = -\tilde{\mathbf{r}}(\mathbf{v})$$

also reads as

$$\nabla_{\mathbf{w}} \mathbf{r}(\mathbf{w})^\top A \Delta\mathbf{v} = -\mathbf{r}(A\mathbf{v} + \mathbf{a}) = -\mathbf{r}(\mathbf{w})$$

Affine invariance of the (exact) Newton method

Affine change of coordinates

Consider: $\mathbf{w} = A\mathbf{v} + \mathbf{a}$ with $A \in \mathbb{R}^{n \times n}$ non-singular and $\mathbf{a} \in \mathbb{R}^n$.

Define $\tilde{\mathbf{r}}(\mathbf{v}) = \mathbf{r}(A\mathbf{v} + \mathbf{a})$ then: $$\nabla_{\mathbf{v}} \tilde{\mathbf{r}}(\mathbf{v}) = A^\top \nabla_{\mathbf{w}} \mathbf{r}(\mathbf{w})$$

The Newton step in $\mathbf{v}$ given by:

$$\nabla_{\mathbf{v}} \tilde{\mathbf{r}}(\mathbf{v})^\top \Delta\mathbf{v} = -\tilde{\mathbf{r}}(\mathbf{v})$$

also reads as

$$\nabla_{\mathbf{w}} \mathbf{r}(\mathbf{w})^\top A \Delta\mathbf{v} = -\mathbf{r}(A\mathbf{v} + \mathbf{a}) = -\mathbf{r}(\mathbf{w})$$

It follows that $\Delta\mathbf{w} = A\Delta\mathbf{v}$ is the Newton step on $\mathbf{r}(\mathbf{w})$:

$$\nabla_{\mathbf{w}} \mathbf{r}(\mathbf{w})^\top \Delta\mathbf{w} = -\mathbf{r}(\mathbf{w})$$

Affine invariance of the (exact) Newton method

Affine change of coordinates

Consider: $\mathbf{w} = A\mathbf{v} + \mathbf{a}$ with $A \in \mathbb{R}^{n \times n}$ non-singular and $\mathbf{a} \in \mathbb{R}^n$.

Define $\tilde{\mathbf{r}}(\mathbf{v}) = \mathbf{r}(A\mathbf{v} + \mathbf{a})$ then: $$\nabla_{\mathbf{v}} \tilde{\mathbf{r}}(\mathbf{v}) = A^\top \nabla_{\mathbf{w}} \mathbf{r}(\mathbf{w})$$

The Newton step in $\mathbf{v}$ given by:

$$\nabla_{\mathbf{v}} \tilde{\mathbf{r}}(\mathbf{v})^\top \Delta\mathbf{v} = -\tilde{\mathbf{r}}(\mathbf{v})$$

also reads as

$$\nabla_{\mathbf{w}} \mathbf{r}(\mathbf{w})^\top A \Delta\mathbf{v} = -\mathbf{r}(A\mathbf{v} + \mathbf{a}) = -\mathbf{r}(\mathbf{w})$$

It follows that $\Delta\mathbf{w} = A\Delta\mathbf{v}$ is the Newton step on $\mathbf{r}(\mathbf{w})$:

$$\nabla_{\mathbf{w}} \mathbf{r}(\mathbf{w})^\top \Delta\mathbf{w} = -\mathbf{r}(\mathbf{w})$$

Original Newton step $\Delta\mathbf{w}$ and "transformed" step $\Delta\mathbf{v}$ are linearly related !!

Affine invariance of the (exact) Newton method

Affine change of coordinates

Consider: $\mathbf{w} = A\mathbf{v} + \mathbf{a}$ with $A \in \mathbb{R}^{n \times n}$ non-singular and $\mathbf{a} \in \mathbb{R}^n$.

Define $\tilde{\mathbf{r}}(\mathbf{v}) = \mathbf{r}(A\mathbf{v} + \mathbf{a})$ then: $$\nabla_{\mathbf{v}} \tilde{\mathbf{r}}(\mathbf{v}) = A^\top \nabla_{\mathbf{w}} \mathbf{r}(\mathbf{w})$$

The Newton step in $\mathbf{v}$ given by:

$$\nabla_{\mathbf{v}} \tilde{\mathbf{r}}(\mathbf{v})^\top \Delta\mathbf{v} = -\tilde{\mathbf{r}}(\mathbf{v})$$

also reads as

$$\nabla_{\mathbf{w}} \mathbf{r}(\mathbf{w})^\top A \Delta\mathbf{v} = -\mathbf{r}(A\mathbf{v} + \mathbf{a}) = -\mathbf{r}(\mathbf{w})$$

It follows that $\Delta\mathbf{w} = A\Delta\mathbf{v}$ is the Newton step on $\mathbf{r}(\mathbf{w})$:

$$\nabla_{\mathbf{w}} \mathbf{r}(\mathbf{w})^\top \Delta\mathbf{w} = -\mathbf{r}(\mathbf{w})$$

Original Newton step $\Delta\mathbf{w}$ and "transformed" step $\Delta\mathbf{v}$ are linearly related !!

- Scaling does not affect the behaviour of the exact Newton method (iterations linearly related). Inexact Newton method can be affected !

Affine invariance of the (exact) Newton method

Affine change of coordinates

Consider: $\mathbf{w} = A\mathbf{v} + \mathbf{a}$ with $A \in \mathbb{R}^{n \times n}$ non-singular and $\mathbf{a} \in \mathbb{R}^n$.

Define $\tilde{\mathbf{r}}(\mathbf{v}) = \mathbf{r}(A\mathbf{v} + \mathbf{a})$ then: $$\nabla_{\mathbf{v}} \tilde{\mathbf{r}}(\mathbf{v}) = A^\top \nabla_{\mathbf{w}} \mathbf{r}(\mathbf{w})$$

The Newton step in $\mathbf{v}$ given by:

$$\nabla_{\mathbf{v}} \tilde{\mathbf{r}}(\mathbf{v})^\top \Delta\mathbf{v} = -\tilde{\mathbf{r}}(\mathbf{v})$$

also reads as

$$\nabla_{\mathbf{w}} \mathbf{r}(\mathbf{w})^\top A \Delta\mathbf{v} = -\mathbf{r}(A\mathbf{v} + \mathbf{a}) = -\mathbf{r}(\mathbf{w})$$

It follows that $\Delta\mathbf{w} = A\Delta\mathbf{v}$ is the Newton step on $\mathbf{r}(\mathbf{w})$:

$$\nabla_{\mathbf{w}} \mathbf{r}(\mathbf{w})^\top \Delta\mathbf{w} = -\mathbf{r}(\mathbf{w})$$

Original Newton step $\Delta\mathbf{w}$ and "transformed" step $\Delta\mathbf{v}$ are linearly related !!

- Scaling does not affect the behaviour of the exact Newton method (iterations linearly related). Inexact Newton method can be affected !
- Convergence proof is not scale-invariant ! Check out self-concordance theory to address that question !

Affine invariance of the (exact) Newton method

Affine change of coordinates

Consider: $\mathbf{w} = A\mathbf{v} + \mathbf{a}$ with $A \in \mathbb{R}^{n \times n}$ non-singular and $\mathbf{a} \in \mathbb{R}^n$.

Define $\tilde{\mathbf{r}}(\mathbf{v}) = \mathbf{r}(A\mathbf{v} + \mathbf{a})$ then: $$\nabla_{\mathbf{v}} \tilde{\mathbf{r}}(\mathbf{v}) = A^\top \nabla_{\mathbf{w}} \mathbf{r}(\mathbf{w})$$

The Newton step in $\mathbf{v}$ given by:

$$\nabla_{\mathbf{v}} \tilde{\mathbf{r}}(\mathbf{v})^\top \Delta\mathbf{v} = -\tilde{\mathbf{r}}(\mathbf{v})$$

also reads as

$$\nabla_{\mathbf{w}} \mathbf{r}(\mathbf{w})^\top A \Delta\mathbf{v} = -\mathbf{r}(A\mathbf{v} + \mathbf{a}) = -\mathbf{r}(\mathbf{w})$$

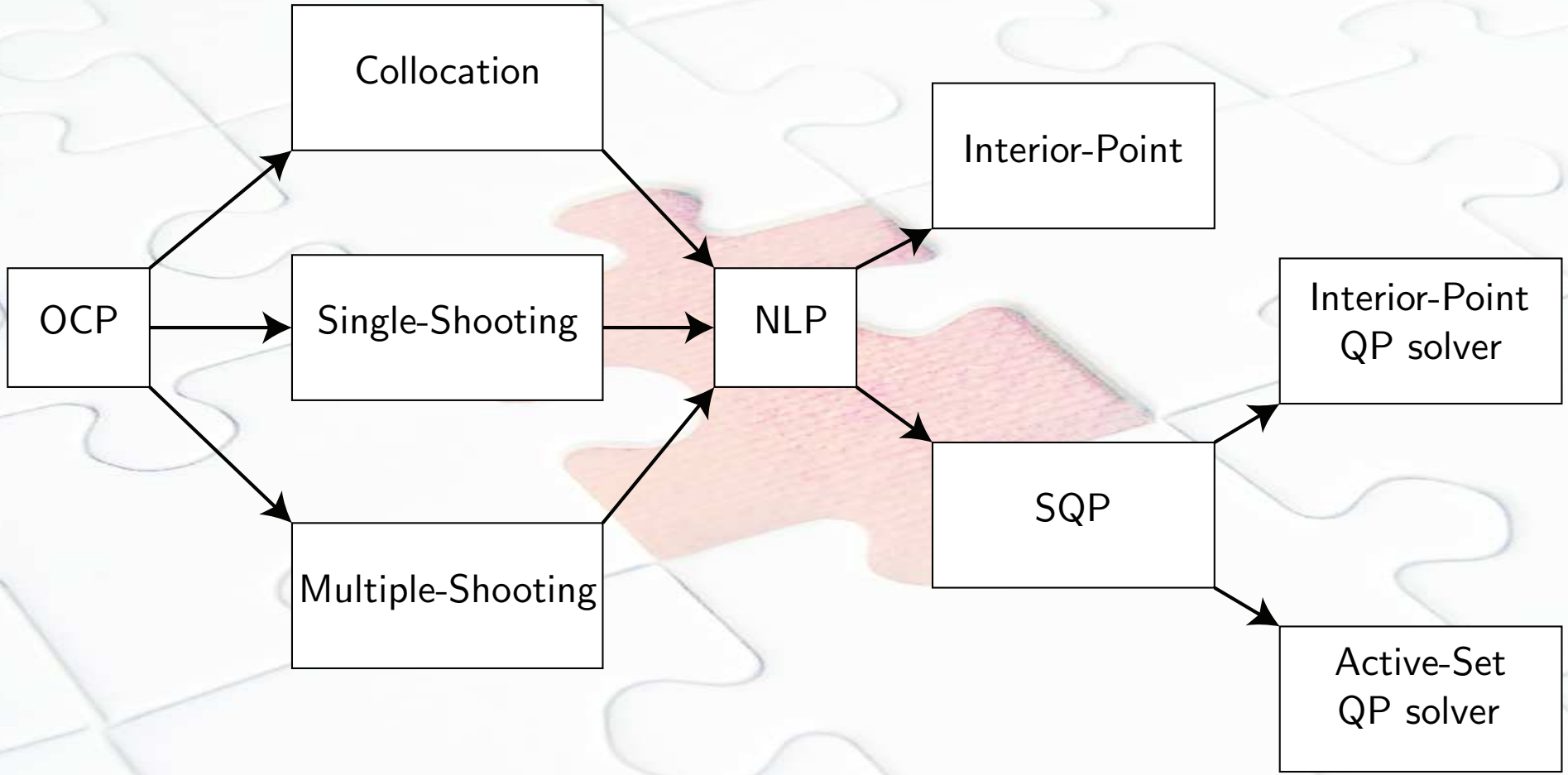
It follows that $\Delta\mathbf{w} = A\Delta\mathbf{v}$ is the Newton step on $\mathbf{r}(\mathbf{w})$:

$$\nabla_{\mathbf{w}} \mathbf{r}(\mathbf{w})^\top \Delta\mathbf{w} = -\mathbf{r}(\mathbf{w})$$

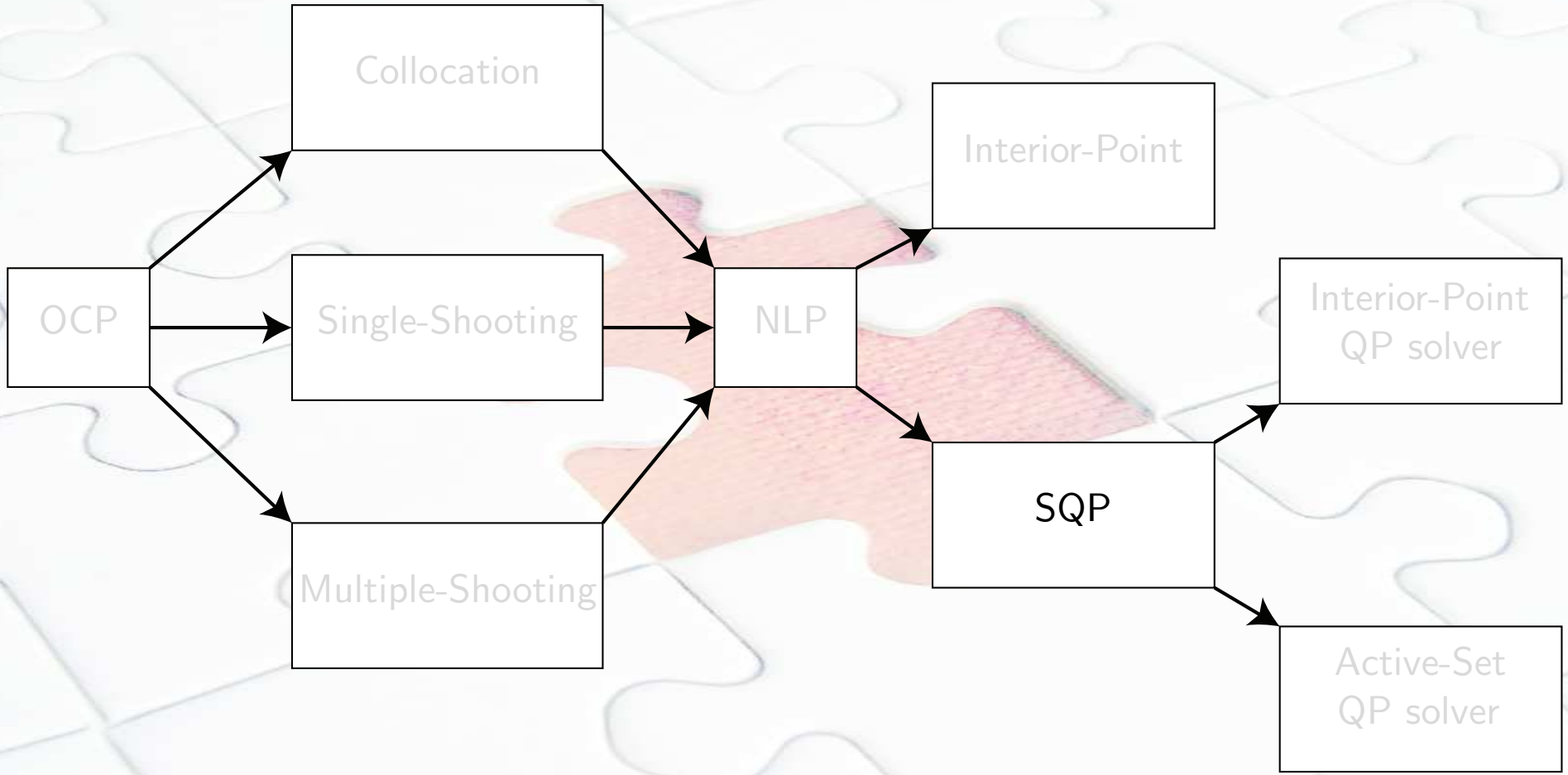
Original Newton step $\Delta\mathbf{w}$ and "transformed" step $\Delta\mathbf{v}$ are linearly related !!

- Scaling does not affect the behaviour of the exact Newton method (iterations linearly related). Inexact Newton method can be affected !
- Convergence proof is not scale-invariant ! Check out self-concordance theory to address that question !
- **Rescaling** can still be very useful to **improve conditioning** of $\nabla_{\mathbf{w}} \mathbf{r}(\mathbf{w})$

Survival map of Direct Optimal Control



Survival map of Direct Optimal Control



Let's approach the problem of solving the KKT conditions

Outline

1. The Newton method
2. Newton on the KKT conditions
3. The reduced Newton step (unconstrained problems)
4. The merit function - Line-search for constrained problems
5. Newton-type methods
6. Sequential Quadratic Programming

Core idea

A vast majority of solvers try to find a point $\mathbf{w}$, $\boldsymbol{\mu}$, $\boldsymbol{\lambda}$ satisfying the KKT conditions:

Primal Feasibility:	$\mathbf{g}(\mathbf{w}) = 0, \quad \mathbf{h}(\mathbf{w}) \leq 0,$
Dual Feasibility:	$\nabla_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \boldsymbol{\mu}, \boldsymbol{\lambda}) = 0, \quad \boldsymbol{\mu} \geq 0,$
Complementarity Slackness:	$\mu_i \mathbf{h}_i(\mathbf{w}) = 0, \quad i = 1, \ldots$

where $\mathcal{L} = \Phi(\mathbf{w}) + \boldsymbol{\lambda}^{\top} \mathbf{g}(\mathbf{w}) + \boldsymbol{\mu}^{\top} \mathbf{h}(\mathbf{w})$

Core idea

A vast majority of solvers try to find a point $\mathbf{w}$, $\boldsymbol{\mu}$, $\boldsymbol{\lambda}$ satisfying the KKT conditions:

Primal Feasibility:	$\mathbf{g}(\mathbf{w}) = 0, \quad \mathbf{h}(\mathbf{w}) \leq 0,$
Dual Feasibility:	$\nabla_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \boldsymbol{\mu}, \boldsymbol{\lambda}) = 0, \quad \boldsymbol{\mu} \geq 0,$
Complementarity Slackness:	$\boldsymbol{\mu}_i \mathbf{h}_i(\mathbf{w}) = 0, \quad i = 1, \ldots$

where $\mathcal{L} = \Phi(\mathbf{w}) + \boldsymbol{\lambda}^\top \mathbf{g}(\mathbf{w}) + \boldsymbol{\mu}^\top \mathbf{h}(\mathbf{w})$

Let's consider *for now* equality constrained problems, i.e. find $\mathbf{w}, \boldsymbol{\lambda}$ s.t.:

$$\begin{aligned} \nabla_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}) &= 0 \\ \mathbf{g}(\mathbf{w}) &= 0 \end{aligned}$$

Core idea

A vast majority of solvers try to find a point $\mathbf{w}$, $\boldsymbol{\mu}$, $\boldsymbol{\lambda}$ satisfying the KKT conditions:

Primal Feasibility:	$\mathbf{g}(\mathbf{w}) = 0, \quad \mathbf{h}(\mathbf{w}) \leq 0,$
Dual Feasibility:	$\nabla_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \boldsymbol{\mu}, \boldsymbol{\lambda}) = 0, \quad \boldsymbol{\mu} \geq 0,$
Complementarity Slackness:	$\mu_i \mathbf{h}_i(\mathbf{w}) = 0, \quad i = 1, \ldots$

where $\mathcal{L} = \Phi(\mathbf{w}) + \boldsymbol{\lambda}^{\top} \mathbf{g}(\mathbf{w}) + \boldsymbol{\mu}^{\top} \mathbf{h}(\mathbf{w})$

Let's consider *for now* equality constrained problems, i.e. find $\mathbf{w}, \boldsymbol{\lambda}$ s.t.:

$$\begin{aligned} \nabla_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}) &= 0 \\ \mathbf{g}(\mathbf{w}) &= 0 \end{aligned}$$

Idea: apply the Newton method on the KKT conditions, i.e.

Solve...

$$\mathbf{r}(\mathbf{w}, \boldsymbol{\lambda}) = \begin{bmatrix} \nabla_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}) \\ \mathbf{g}(\mathbf{w}) \end{bmatrix} = 0$$

Core idea

A vast majority of solvers try to find a point $\mathbf{w}$, $\boldsymbol{\mu}$, $\boldsymbol{\lambda}$ satisfying the KKT conditions:

Primal Feasibility:	$\mathbf{g}(\mathbf{w}) = 0, \quad \mathbf{h}(\mathbf{w}) \leq 0,$
Dual Feasibility:	$\nabla_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \boldsymbol{\mu}, \boldsymbol{\lambda}) = 0, \quad \boldsymbol{\mu} \geq 0,$
Complementarity Slackness:	$\mu_i \mathbf{h}_i(\mathbf{w}) = 0, \quad i = 1, \ldots$

where $\mathcal{L} = \Phi(\mathbf{w}) + \boldsymbol{\lambda}^{\top} \mathbf{g}(\mathbf{w}) + \boldsymbol{\mu}^{\top} \mathbf{h}(\mathbf{w})$

Let's consider *for now* equality constrained problems, i.e. find $\mathbf{w}, \boldsymbol{\lambda}$ s.t.:

$$\begin{aligned} \nabla_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}) &= 0 \\ \mathbf{g}(\mathbf{w}) &= 0 \end{aligned}$$

Idea: apply the Newton method on the KKT conditions, i.e.

Solve...

$$\mathbf{r}(\mathbf{w}, \boldsymbol{\lambda}) = \begin{bmatrix} \nabla_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}) \\ \mathbf{g}(\mathbf{w}) \end{bmatrix} = 0$$

... by iterating

$$\nabla \mathbf{r}(\mathbf{w}, \boldsymbol{\lambda})^{\top} \begin{bmatrix} \Delta \mathbf{w} \\ \Delta \boldsymbol{\lambda} \end{bmatrix} = -\mathbf{r}(\mathbf{w}, \boldsymbol{\lambda})$$

Newton method on the KKT conditions

KKT conditions:

$$\nabla_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}) = 0$$
$$\mathbf{g}(\mathbf{w}) = 0$$

Newton iteration:

$$\nabla \mathbf{r}(\mathbf{w}, \boldsymbol{\lambda})^{\top} \begin{bmatrix} \Delta \mathbf{w} \\ \Delta \boldsymbol{\lambda} \end{bmatrix} = -\mathbf{r}(\mathbf{w}, \boldsymbol{\lambda})$$

Newton direction:

Newton method on the KKT conditions

KKT conditions:

$$\begin{aligned} \nabla_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}) &= 0 \\ \mathbf{g}(\mathbf{w}) &= 0 \end{aligned}$$

Newton iteration:

$$\nabla \mathbf{r}\left(\mathbf{w}, \boldsymbol{\lambda}\right)^{\top} \begin{bmatrix} \Delta \mathbf{w} \\ \Delta \boldsymbol{\lambda} \end{bmatrix} = -\mathbf{r}\left(\mathbf{w}, \boldsymbol{\lambda}\right)$$

Newton direction:

$$\begin{array}{rcll} \nabla_{\mathbf{w}}^2 \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}) \Delta \mathbf{w} & + & \nabla_{\mathbf{w}, \boldsymbol{\lambda}} \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}) \Delta \boldsymbol{\lambda} & = & -\nabla_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}) \\ \nabla \mathbf{g}(\mathbf{w})^{\mathsf{T}} \Delta \mathbf{w} & & & = & -\mathbf{g}(\mathbf{w}) \end{array}$$

Newton method on the KKT conditions

KKT conditions:

$$\begin{aligned} \nabla_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}) &= 0 \\ \mathbf{g}(\mathbf{w}) &= 0 \end{aligned}$$

Newton iteration:

$$\nabla \mathbf{r}\left(\mathbf{w}, \boldsymbol{\lambda}\right)^{\top} \begin{bmatrix} \Delta \mathbf{w} \\ \Delta \boldsymbol{\lambda} \end{bmatrix} = -\mathbf{r}\left(\mathbf{w}, \boldsymbol{\lambda}\right)$$

Newton direction: using $\nabla_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}) = \nabla \Phi(\mathbf{w}) + \nabla \mathbf{g}(\mathbf{w}) \boldsymbol{\lambda}$

$$\begin{array}{rcrcl} \nabla_{\mathbf{w}}^2 \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}) \Delta \mathbf{w} & + & \nabla_{\mathbf{w}, \boldsymbol{\lambda}} \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}) \Delta \boldsymbol{\lambda} & = & -\nabla_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}) \\ \nabla \mathbf{g}(\mathbf{w})^{\mathsf{T}} \Delta \mathbf{w} & & & = & -\mathbf{g}(\mathbf{w}) \end{array}$$

Newton method on the KKT conditions

KKT conditions:

$$\begin{aligned} \nabla_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}) &= 0 \\ \mathbf{g}(\mathbf{w}) &= 0 \end{aligned}$$

Newton iteration:

$$\nabla \mathbf{r}\left(\mathbf{w}, \boldsymbol{\lambda}\right)^{\top} \begin{bmatrix} \Delta \mathbf{w} \\ \Delta \boldsymbol{\lambda} \end{bmatrix} = -\mathbf{r}\left(\mathbf{w}, \boldsymbol{\lambda}\right)$$

Newton direction: using $\nabla_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}) = \nabla \Phi(\mathbf{w}) + \nabla \mathbf{g}(\mathbf{w}) \boldsymbol{\lambda}$

$$\begin{array}{rcrcl} \nabla_{\mathbf{w}}^2 \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}) \Delta \mathbf{w} & + & \nabla \mathbf{g}(\mathbf{w}) \Delta \boldsymbol{\lambda} & = & -\nabla_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}) \\ \nabla \mathbf{g}(\mathbf{w})^{\mathsf{T}} \Delta \mathbf{w} & & & = & -\mathbf{g}(\mathbf{w}) \end{array}$$

Newton method on the KKT conditions

KKT conditions:

$$\begin{aligned} \nabla_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}) &= 0 \\ \mathbf{g}(\mathbf{w}) &= 0 \end{aligned}$$

Newton iteration:

$$\nabla \mathbf{r}\left(\mathbf{w}, \boldsymbol{\lambda}\right)^{\top} \begin{bmatrix} \Delta \mathbf{w} \\ \Delta \boldsymbol{\lambda} \end{bmatrix} = -\mathbf{r}\left(\mathbf{w}, \boldsymbol{\lambda}\right)$$

Newton direction: using $\nabla_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}) = \nabla \Phi(\mathbf{w}) + \nabla \mathbf{g}(\mathbf{w}) \boldsymbol{\lambda}$

$$\begin{array}{rcrcl} \nabla_{\mathbf{w}}^2 \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}) \Delta \mathbf{w} & + & \nabla \mathbf{g}(\mathbf{w}) \Delta \boldsymbol{\lambda} & = & -\nabla \Phi(\mathbf{w}) - \nabla \mathbf{g}(\mathbf{w}) \boldsymbol{\lambda} \\ \nabla \mathbf{g}(\mathbf{w})^{\mathsf{T}} \Delta \mathbf{w} & & & = & -\mathbf{g}(\mathbf{w}) \end{array}$$

Newton method on the KKT conditions

KKT conditions:

$$\begin{aligned} \nabla_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}) &= 0 \\ \mathbf{g}(\mathbf{w}) &= 0 \end{aligned}$$

Newton iteration:

$$\nabla \mathbf{r}\left(\mathbf{w}, \boldsymbol{\lambda}\right)^{\top} \begin{bmatrix} \Delta \mathbf{w} \\ \Delta \boldsymbol{\lambda} \end{bmatrix} = -\mathbf{r}\left(\mathbf{w}, \boldsymbol{\lambda}\right)$$

Newton direction: using $\nabla_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}) = \nabla \Phi(\mathbf{w}) + \nabla \mathbf{g}(\mathbf{w}) \boldsymbol{\lambda}$

$$\begin{array}{rcll} \nabla_{\mathbf{w}}^2 \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}) \Delta \mathbf{w} & + & \nabla \mathbf{g}(\mathbf{w}) \left(\boldsymbol{\lambda} + \Delta \boldsymbol{\lambda}\right) & = & -\nabla \Phi(\mathbf{w}) \\ \nabla \mathbf{g}(\mathbf{w})^{\mathsf{T}} \Delta \mathbf{w} & & & = & -\mathbf{g}(\mathbf{w}) \end{array}$$

Newton method on the KKT conditions

KKT conditions:

$$\begin{aligned} \nabla_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}) &= 0 \\ \mathbf{g}(\mathbf{w}) &= 0 \end{aligned}$$

Newton iteration:

$$\nabla \mathbf{r}(\mathbf{w}, \boldsymbol{\lambda})^{\top} \begin{bmatrix} \Delta \mathbf{w} \\ \Delta \boldsymbol{\lambda} \end{bmatrix} = -\mathbf{r}(\mathbf{w}, \boldsymbol{\lambda})$$

Newton direction: using $\nabla_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}) = \nabla \Phi(\mathbf{w}) + \nabla \mathbf{g}(\mathbf{w}) \boldsymbol{\lambda}$

$$\begin{array}{rcl} \nabla_{\mathbf{w}}^2 \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}) \Delta \mathbf{w} + \nabla \mathbf{g}(\mathbf{w}) (\boldsymbol{\lambda} + \Delta \boldsymbol{\lambda}) &=& -\nabla \Phi(\mathbf{w}) \\ \nabla \mathbf{g}(\mathbf{w})^{\mathsf{T}} \Delta \mathbf{w} &=& -\mathbf{g}(\mathbf{w}) \end{array}$$

The Newton direction on the KKT conditions

$$\begin{bmatrix} \nabla_{\mathbf{w}}^2 \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}) & \nabla \mathbf{g}(\mathbf{w}) \\ \nabla \mathbf{g}(\mathbf{w})^{\mathsf{T}} & 0 \end{bmatrix} \begin{bmatrix} \Delta \mathbf{w} \\ \boldsymbol{\lambda} + \Delta \boldsymbol{\lambda} \end{bmatrix} = - \begin{bmatrix} \nabla \Phi(\mathbf{w}) \\ \mathbf{g}(\mathbf{w}) \end{bmatrix}$$

Newton method on the KKT conditions

KKT conditions:

$$\begin{aligned} \nabla_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}) &= 0 \\ \mathbf{g}(\mathbf{w}) &= 0 \end{aligned}$$

Newton iteration:

$$\nabla \mathbf{r}(\mathbf{w}, \boldsymbol{\lambda})^{\top} \begin{bmatrix} \Delta \mathbf{w} \\ \Delta \boldsymbol{\lambda} \end{bmatrix} = -\mathbf{r}(\mathbf{w}, \boldsymbol{\lambda})$$

Newton direction: using $\nabla_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}) = \nabla \Phi(\mathbf{w}) + \nabla \mathbf{g}(\mathbf{w}) \boldsymbol{\lambda}$

$$\begin{array}{rcrcl} \nabla^2_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}) \Delta \mathbf{w} & + & \nabla \mathbf{g}(\mathbf{w}) (\boldsymbol{\lambda} + \Delta \boldsymbol{\lambda}) & = & -\nabla \Phi(\mathbf{w}) \\ \nabla \mathbf{g}(\mathbf{w})^{\mathsf{T}} \Delta \mathbf{w} & & & = & -\mathbf{g}(\mathbf{w}) \end{array}$$

The Newton direction on the KKT conditions

$$\underbrace{\begin{bmatrix} H(\mathbf{w}, \boldsymbol{\lambda}) & \nabla \mathbf{g}(\mathbf{w}) \\ \nabla \mathbf{g}(\mathbf{w})^{\mathsf{T}} & 0 \end{bmatrix}}_{\text{KKT matrix (symmetric indefinite)}} \begin{bmatrix} \Delta \mathbf{w} \\ \boldsymbol{\lambda} + \Delta \boldsymbol{\lambda} \end{bmatrix} = - \begin{bmatrix} \nabla \Phi(\mathbf{w}) \\ \mathbf{g}(\mathbf{w}) \end{bmatrix}$$

where $H(\mathbf{w}, \boldsymbol{\lambda}) = \nabla^2_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda})$ is the Hessian of the problem.

Newton method on the KKT conditions

KKT conditions:

$$\begin{aligned} \nabla_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}) &= 0 \\ \mathbf{g}(\mathbf{w}) &= 0 \end{aligned}$$

Newton iteration:

$$\nabla \mathbf{r}(\mathbf{w}, \boldsymbol{\lambda})^{\top} \begin{bmatrix} \Delta \mathbf{w} \\ \Delta \boldsymbol{\lambda} \end{bmatrix} = -\mathbf{r}(\mathbf{w}, \boldsymbol{\lambda})$$

Newton direction: using $\nabla_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}) = \nabla \Phi(\mathbf{w}) + \nabla \mathbf{g}(\mathbf{w}) \boldsymbol{\lambda}$

$$\begin{array}{rcrcl} \nabla^2_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}) \Delta \mathbf{w} & + & \nabla \mathbf{g}(\mathbf{w}) (\boldsymbol{\lambda} + \Delta \boldsymbol{\lambda}) & = & -\nabla \Phi(\mathbf{w}) \\ \nabla \mathbf{g}(\mathbf{w})^{\mathsf{T}} \Delta \mathbf{w} & & & = & -\mathbf{g}(\mathbf{w}) \end{array}$$

The Newton direction on the KKT conditions

$$\underbrace{\begin{bmatrix} H(\mathbf{w}, \boldsymbol{\lambda}) & \nabla \mathbf{g}(\mathbf{w}) \\ \nabla \mathbf{g}(\mathbf{w})^{\mathsf{T}} & 0 \end{bmatrix}}_{\text{KKT matrix (symmetric indefinite)}} \begin{bmatrix} \Delta \mathbf{w} \\ \boldsymbol{\lambda}^{+} \end{bmatrix} = - \begin{bmatrix} \nabla \Phi(\mathbf{w}) \\ \mathbf{g}(\mathbf{w}) \end{bmatrix}$$

where $H(\mathbf{w}, \boldsymbol{\lambda}) = \nabla^2_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda})$ is the Hessian of the problem.

- In the following, we label $\boldsymbol{\lambda}^{+} = \boldsymbol{\lambda} + \Delta \boldsymbol{\lambda}$ (full dual Newton step)

Newton method on the KKT conditions

KKT conditions:

$$\begin{aligned} \nabla_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}) &= 0 \\ \mathbf{g}(\mathbf{w}) &= 0 \end{aligned}$$

Newton iteration:

$$\nabla \mathbf{r}(\mathbf{w}, \boldsymbol{\lambda})^{\top} \begin{bmatrix} \Delta \mathbf{w} \\ \Delta \boldsymbol{\lambda} \end{bmatrix} = -\mathbf{r}(\mathbf{w}, \boldsymbol{\lambda})$$

Newton direction: using $\nabla_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}) = \nabla \Phi(\mathbf{w}) + \nabla \mathbf{g}(\mathbf{w}) \boldsymbol{\lambda}$

$$\begin{array}{rcrcl} \nabla^2_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}) \Delta \mathbf{w} & + & \nabla \mathbf{g}(\mathbf{w}) (\boldsymbol{\lambda} + \Delta \boldsymbol{\lambda}) & = & -\nabla \Phi(\mathbf{w}) \\ \nabla \mathbf{g}(\mathbf{w})^{\mathsf{T}} \Delta \mathbf{w} & & & = & -\mathbf{g}(\mathbf{w}) \end{array}$$

The Newton direction on the KKT conditions

$$\underbrace{\begin{bmatrix} H(\mathbf{w}, \boldsymbol{\lambda}) & \nabla \mathbf{g}(\mathbf{w}) \\ \nabla \mathbf{g}(\mathbf{w})^{\mathsf{T}} & 0 \end{bmatrix}}_{\text{KKT matrix (symmetric indefinite)}} \begin{bmatrix} \Delta \mathbf{w} \\ \boldsymbol{\lambda}^{+} \end{bmatrix} = - \begin{bmatrix} \nabla \Phi(\mathbf{w}) \\ \mathbf{g}(\mathbf{w}) \end{bmatrix}$$

where $H(\mathbf{w}, \boldsymbol{\lambda}) = \nabla^2_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda})$ is the Hessian of the problem.

- In the following, we label $\boldsymbol{\lambda}^{+} = \boldsymbol{\lambda} + \Delta \boldsymbol{\lambda}$ (full dual Newton step)
- $\mathcal{L}(\mathbf{w}, \boldsymbol{\lambda})$ enters only in the Hessian !

Newton method on the KKT conditions

KKT conditions:

$$\begin{aligned} \nabla_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}) &= 0 \\ \mathbf{g}(\mathbf{w}) &= 0 \end{aligned}$$

Newton iteration:

$$\nabla \mathbf{r}(\mathbf{w}, \boldsymbol{\lambda})^{\top} \begin{bmatrix} \Delta \mathbf{w} \\ \Delta \boldsymbol{\lambda} \end{bmatrix} = -\mathbf{r}(\mathbf{w}, \boldsymbol{\lambda})$$

Newton direction: using $\nabla_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}) = \nabla \Phi(\mathbf{w}) + \nabla \mathbf{g}(\mathbf{w}) \boldsymbol{\lambda}$

$$\begin{array}{rcrcl} \nabla^2_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}) \Delta \mathbf{w} & + & \nabla \mathbf{g}(\mathbf{w}) (\boldsymbol{\lambda} + \Delta \boldsymbol{\lambda}) & = & -\nabla \Phi(\mathbf{w}) \\ \nabla \mathbf{g}(\mathbf{w})^{\mathsf{T}} \Delta \mathbf{w} & & & = & -\mathbf{g}(\mathbf{w}) \end{array}$$

The Newton direction on the KKT conditions

$$\underbrace{\begin{bmatrix} H(\mathbf{w}, \boldsymbol{\lambda}) & \nabla \mathbf{g}(\mathbf{w}) \\ \nabla \mathbf{g}(\mathbf{w})^{\mathsf{T}} & 0 \end{bmatrix}}_{\text{KKT matrix (symmetric indefinite)}} \begin{bmatrix} \Delta \mathbf{w} \\ \boldsymbol{\lambda}^{+} \end{bmatrix} = - \begin{bmatrix} \nabla \Phi(\mathbf{w}) \\ \mathbf{g}(\mathbf{w}) \end{bmatrix}$$

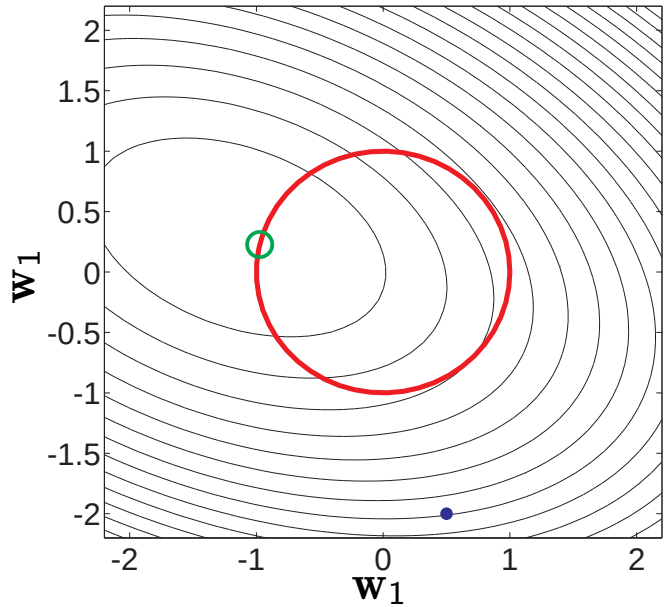
where $H(\mathbf{w}, \boldsymbol{\lambda}) = \nabla^2_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda})$ is the Hessian of the problem.

- In the following, we label $\boldsymbol{\lambda}^{+} = \boldsymbol{\lambda} + \Delta \boldsymbol{\lambda}$ (full dual Newton step)
- $\mathcal{L}(\mathbf{w}, \boldsymbol{\lambda})$ enters only in the Hessian !
- The updated multiplier $\boldsymbol{\lambda}^{+}$ is readily provided...

Newton Iteration for Optimization - Example

$$\min_{\mathbf{w}} \frac{1}{2}\mathbf{w}^T \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix} \mathbf{w} + \mathbf{w}^T \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\text{s.t. } g(\mathbf{w}) = \mathbf{w}^T\mathbf{w} - 1 = 0$$



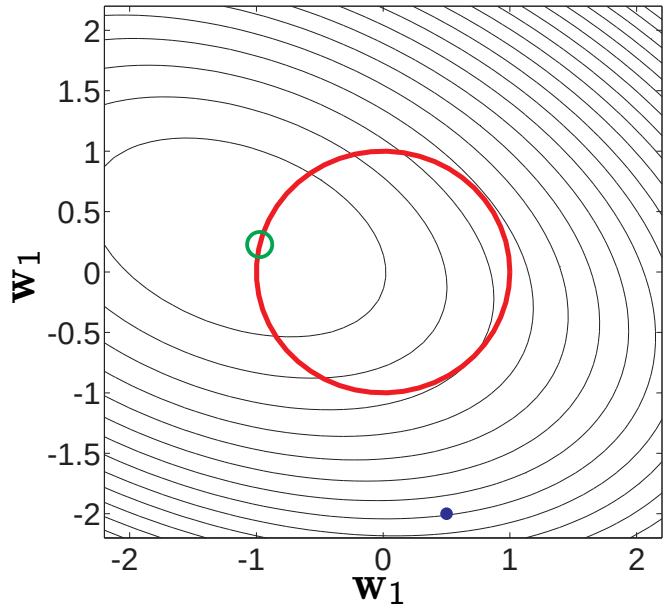
Newton Iteration for Optimization - Example

Iterate:

$$\begin{bmatrix} H & \nabla \mathbf{g} \\ \nabla \mathbf{g}^{\mathsf{T}} & 0 \end{bmatrix} \begin{bmatrix} \Delta \mathbf{w} \\ \boldsymbol{\lambda}^{+} \end{bmatrix} = - \begin{bmatrix} \nabla \Phi \\ \mathbf{g} \end{bmatrix}$$

$$\min_{\mathbf{w}} \frac{1}{2} \mathbf{w}^{\mathsf{T}} \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix} \mathbf{w} + \mathbf{w}^{\mathsf{T}} \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\text{s.t. } g(\mathbf{w}) = \mathbf{w}^{\mathsf{T}} \mathbf{w} - 1 = 0$$



Newton Iteration for Optimization - Example

Iterate:

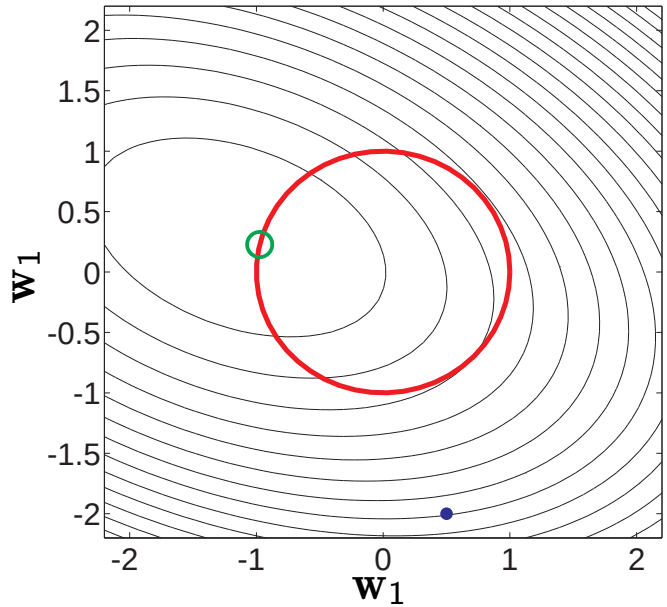
$$\begin{bmatrix} H & \nabla \mathbf{g} \\ \nabla \mathbf{g}^{\mathsf{T}} & 0 \end{bmatrix} \begin{bmatrix} \Delta \mathbf{w} \\ \boldsymbol{\lambda}^{+} \end{bmatrix} = - \begin{bmatrix} \nabla \Phi \\ \mathbf{g} \end{bmatrix}$$

with:

$$\nabla g(\mathbf{w}) = 2\mathbf{w} = \begin{bmatrix} 2\mathbf{w}_1 \\ 2\mathbf{w}_2 \end{bmatrix}$$

$$\min_{\mathbf{w}} \frac{1}{2}\mathbf{w}^{\mathsf{T}} \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix} \mathbf{w} + \mathbf{w}^{\mathsf{T}} \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\text{s.t. } g(\mathbf{w}) = \mathbf{w}^{\mathsf{T}}\mathbf{w} - 1 = 0$$



Newton Iteration for Optimization - Example

Iterate:

$$\begin{bmatrix} H & \nabla \mathbf{g} \\ \nabla \mathbf{g}^{\mathsf{T}} & 0 \end{bmatrix} \begin{bmatrix} \Delta \mathbf{w} \\ \boldsymbol{\lambda}^{+} \end{bmatrix} = - \begin{bmatrix} \nabla \Phi \\ \mathbf{g} \end{bmatrix}$$

with:

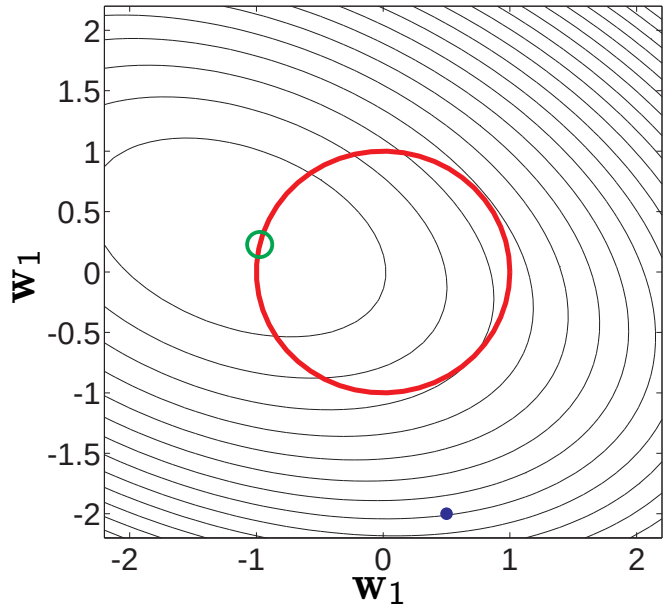
$$\nabla g(\mathbf{w}) = 2\mathbf{w} = \begin{bmatrix} 2\mathbf{w}_1 \\ 2\mathbf{w}_2 \end{bmatrix}$$

$$\mathcal{L}(\mathbf{w}, \lambda) = \Phi(\mathbf{w}) + \lambda g(\mathbf{w})$$

$$\nabla_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \lambda) = \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix} \mathbf{w} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 2\lambda \mathbf{w}$$

$$\min_{\mathbf{w}} \frac{1}{2} \mathbf{w}^{\mathsf{T}} \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix} \mathbf{w} + \mathbf{w}^{\mathsf{T}} \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\text{s.t. } g(\mathbf{w}) = \mathbf{w}^{\mathsf{T}} \mathbf{w} - 1 = 0$$



Newton Iteration for Optimization - Example

Iterate:

$$\begin{bmatrix} H & \nabla g \\ \nabla g^T & 0 \end{bmatrix} \begin{bmatrix} \Delta w \\ \lambda^+ \end{bmatrix} = - \begin{bmatrix} \nabla \Phi \\ g \end{bmatrix}$$

with:

$$\nabla g(w) = 2w = \begin{bmatrix} 2w_1 \\ 2w_2 \end{bmatrix}$$

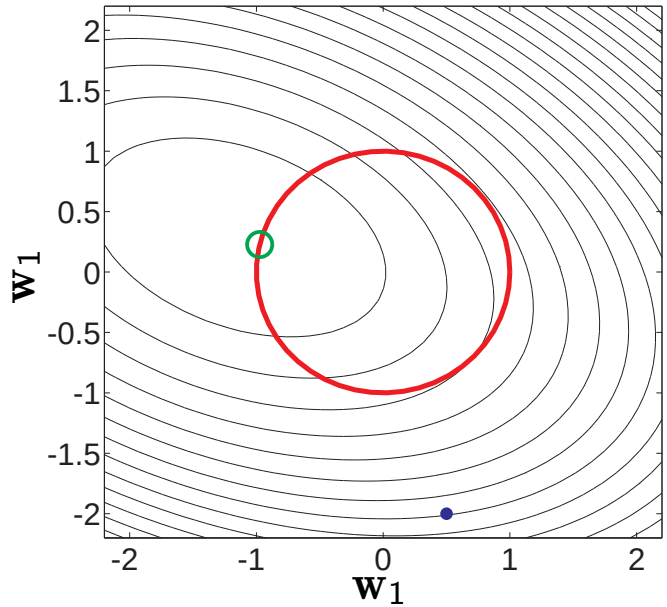
$$\mathcal{L}(w, \lambda) = \Phi(w) + \lambda g(w)$$

$$\nabla_w \mathcal{L}(w, \lambda) = \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix} w + \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 2\lambda w$$

$$H(w, \lambda) = \begin{bmatrix} 2 + 2\lambda & 1 \\ 1 & 4 + 2\lambda \end{bmatrix}$$

$$\min_w \frac{1}{2} w^T \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix} w + w^T \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\text{s.t. } g(w) = w^T w - 1 = 0$$



Newton Iteration for Optimization - Example

Iterate:

$$\begin{bmatrix} H & \nabla g \\ \nabla g^T & 0 \end{bmatrix} \begin{bmatrix} \Delta w \\ \lambda^+ \end{bmatrix} = - \begin{bmatrix} \nabla \Phi \\ g \end{bmatrix}$$

with:

$$\nabla g(w) = 2w = \begin{bmatrix} 2w_1 \\ 2w_2 \end{bmatrix}$$

$$\mathcal{L}(w, \lambda) = \Phi(w) + \lambda g(w)$$

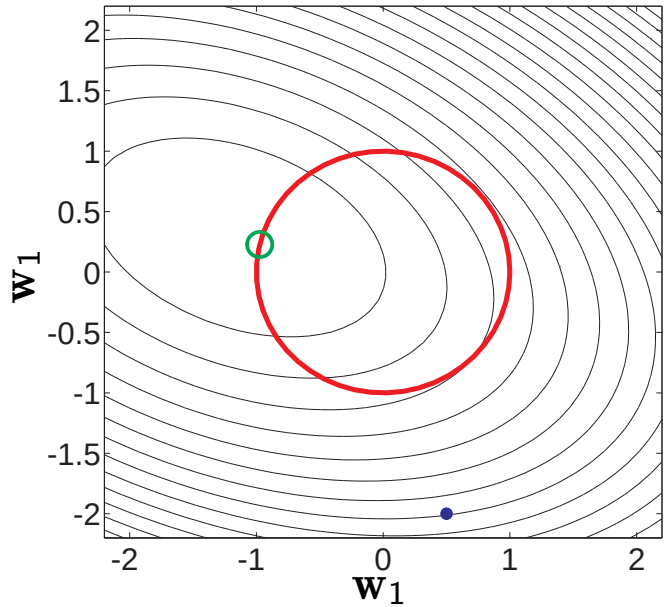
$$\nabla_w \mathcal{L}(w, \lambda) = \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix} w + \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 2\lambda w$$

$$H(w, \lambda) = \begin{bmatrix} 2 + 2\lambda & 1 \\ 1 & 4 + 2\lambda \end{bmatrix}$$

$$\nabla \Phi(w) = \begin{bmatrix} 2w_1 + w_2 + 1 \\ w_1 + 4w_2 \end{bmatrix}$$

$$\min_w \frac{1}{2} w^T \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix} w + w^T \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\text{s.t. } g(w) = w^T w - 1 = 0$$



Newton Iteration for Optimization - Example

Algorithm: Newton method

Input: guess $\mathbf{w}$, $\boldsymbol{\lambda}$

while $\|\nabla \mathcal{L}\|$ *or* $\|\mathbf{g}\| \geq \text{tol}$ **do**

Compute

$$H(\mathbf{w}, \boldsymbol{\lambda}),\ \nabla \mathbf{g}(\mathbf{w}),\ \nabla \Phi(\mathbf{w}),\ \mathbf{g}(\mathbf{w})$$

Compute Newton direction

$$\begin{bmatrix} H & \nabla \mathbf{g} \\ \nabla \mathbf{g}^T & 0 \end{bmatrix} \begin{bmatrix} \Delta \mathbf{w} \\ \boldsymbol{\lambda}^+ \end{bmatrix} = - \begin{bmatrix} \nabla \Phi \\ \mathbf{g} \end{bmatrix}$$

$$\Delta \boldsymbol{\lambda} = \boldsymbol{\lambda}^+ - \boldsymbol{\lambda}$$

Compute Newton step, $t \in]0, 1]$

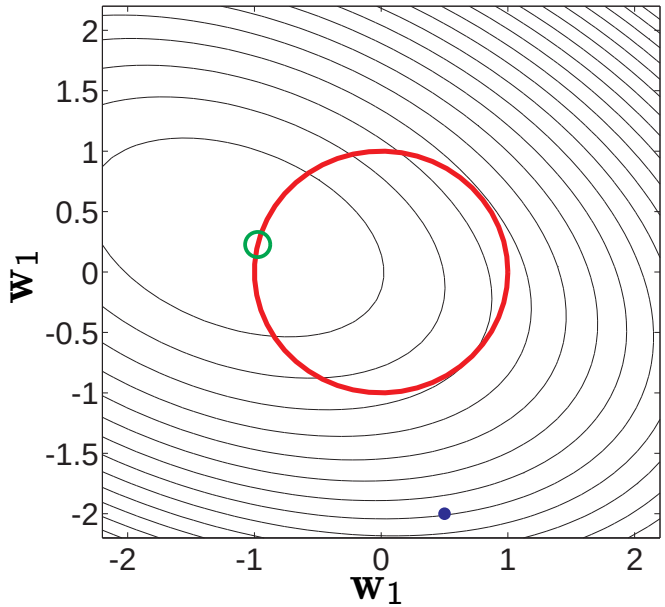
$$\mathbf{w} \leftarrow \mathbf{w} + t \Delta \mathbf{w}, \quad \boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda} + t \Delta \boldsymbol{\lambda}$$

return $\mathbf{w}$, $\boldsymbol{\lambda}$

$$\min_{\mathbf{w}} \ \frac{1}{2} \mathbf{w}^T \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix} \mathbf{w} + \mathbf{w}^T \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\text{s.t. } g(\mathbf{w}) = \mathbf{w}^T \mathbf{w} - 1 = 0$$

Guess $\lambda = 0$, step $t = 1$



Newton Iteration for Optimization - Example

Algorithm: Newton method

Input: guess $\mathbf{w}$, $\boldsymbol{\lambda}$

while $\|\nabla \mathcal{L}\|$ *or* $\|\mathbf{g}\| \geq \text{tol}$ **do**

Compute

$$H(\mathbf{w}, \boldsymbol{\lambda}),\ \nabla \mathbf{g}(\mathbf{w}),\ \nabla \Phi(\mathbf{w}),\ \mathbf{g}(\mathbf{w})$$

Compute Newton direction

$$\begin{bmatrix} H & \nabla \mathbf{g} \\ \nabla \mathbf{g}^T & 0 \end{bmatrix} \begin{bmatrix} \Delta \mathbf{w} \\ \boldsymbol{\lambda}^+ \end{bmatrix} = - \begin{bmatrix} \nabla \Phi \\ \mathbf{g} \end{bmatrix}$$

$$\Delta \boldsymbol{\lambda} = \boldsymbol{\lambda}^+ - \boldsymbol{\lambda}$$

Compute Newton step, $t \in]0, 1]$

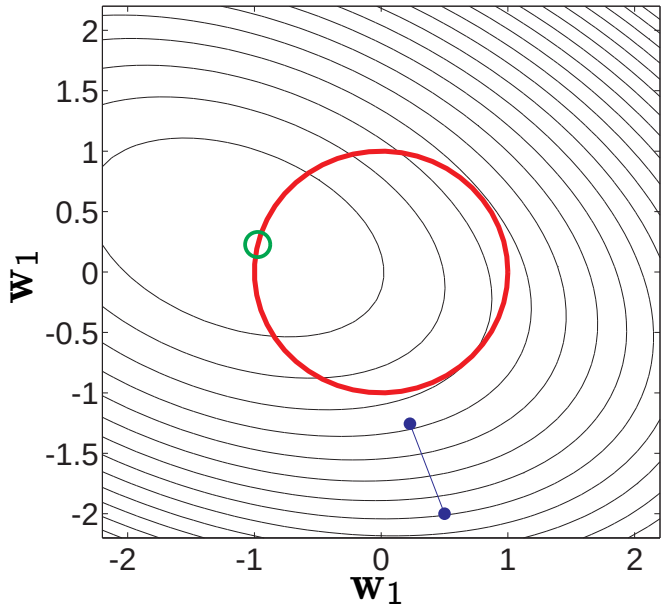
$$\mathbf{w} \leftarrow \mathbf{w} + t \Delta \mathbf{w}, \quad \boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda} + t \Delta \boldsymbol{\lambda}$$

return $\mathbf{w}$, $\boldsymbol{\lambda}$

$$\min_{\mathbf{w}} \ \frac{1}{2} \mathbf{w}^T \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix} \mathbf{w} + \mathbf{w}^T \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\text{s.t. } g(\mathbf{w}) = \mathbf{w}^T \mathbf{w} - 1 = 0$$

Guess $\lambda = 0$, step $t = 1$



Newton Iteration for Optimization - Example

Algorithm: Newton method

Input: guess $\mathbf{w}$, $\boldsymbol{\lambda}$

while $\|\nabla \mathcal{L}\|$ *or* $\|\mathbf{g}\| \geq \text{tol}$ **do**

Compute

$$H(\mathbf{w}, \boldsymbol{\lambda}),\ \nabla \mathbf{g}(\mathbf{w}),\ \nabla \Phi(\mathbf{w}),\ \mathbf{g}(\mathbf{w})$$

Compute Newton direction

$$\begin{bmatrix} H & \nabla \mathbf{g} \\ \nabla \mathbf{g}^T & 0 \end{bmatrix} \begin{bmatrix} \Delta \mathbf{w} \\ \boldsymbol{\lambda}^+ \end{bmatrix} = - \begin{bmatrix} \nabla \Phi \\ \mathbf{g} \end{bmatrix}$$

$$\Delta \boldsymbol{\lambda} = \boldsymbol{\lambda}^+ - \boldsymbol{\lambda}$$

Compute Newton step, $t \in]0, 1]$

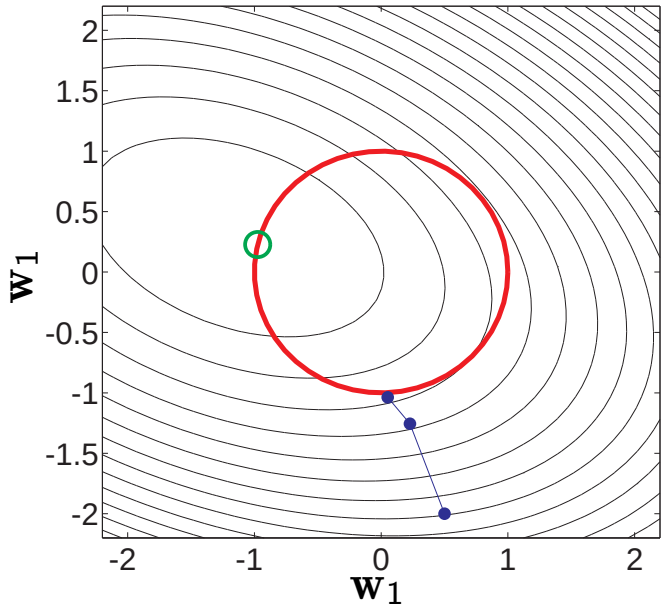
$$\mathbf{w} \leftarrow \mathbf{w} + t \Delta \mathbf{w}, \quad \boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda} + t \Delta \boldsymbol{\lambda}$$

return $\mathbf{w}$, $\boldsymbol{\lambda}$

$$\min_{\mathbf{w}} \ \frac{1}{2} \mathbf{w}^T \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix} \mathbf{w} + \mathbf{w}^T \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\text{s.t. } g(\mathbf{w}) = \mathbf{w}^T \mathbf{w} - 1 = 0$$

Guess $\lambda = 0$, step $t = 1$



Newton Iteration for Optimization - Example

Algorithm: Newton method

Input: guess $\mathbf{w}$, $\boldsymbol{\lambda}$

while $\|\nabla \mathcal{L}\|$ *or* $\|\mathbf{g}\| \geq \mathrm{tol}$ **do**

Compute

$$H(\mathbf{w}, \boldsymbol{\lambda}),\ \nabla \mathbf{g}(\mathbf{w}),\ \nabla \Phi(\mathbf{w}),\ \mathbf{g}(\mathbf{w})$$

Compute Newton direction

$$\begin{bmatrix} H & \nabla \mathbf{g} \\ \nabla \mathbf{g}^{\mathsf{T}} & 0 \end{bmatrix} \begin{bmatrix} \Delta \mathbf{w} \\ \boldsymbol{\lambda}^{+} \end{bmatrix} = - \begin{bmatrix} \nabla \Phi \\ \mathbf{g} \end{bmatrix}$$

$$\Delta \boldsymbol{\lambda} = \boldsymbol{\lambda}^{+} - \boldsymbol{\lambda}$$

Compute Newton step, $t \in]0, 1]$

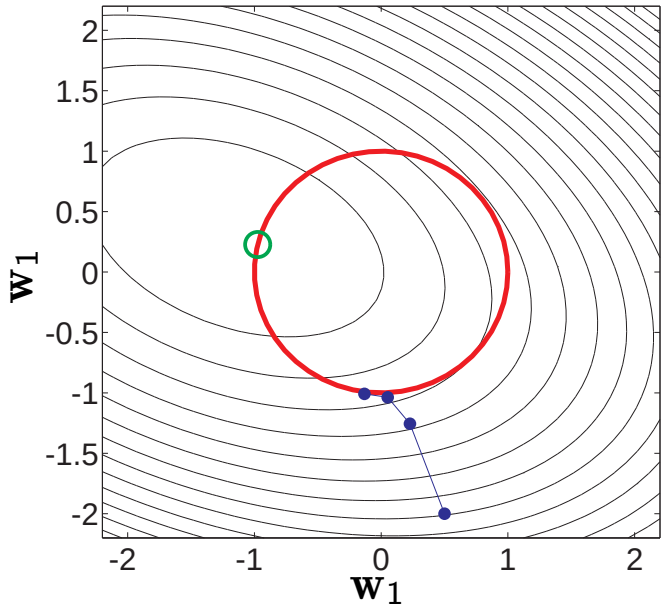
$$\mathbf{w} \leftarrow \mathbf{w} + t \Delta \mathbf{w}, \quad \boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda} + t \Delta \boldsymbol{\lambda}$$

return $\mathbf{w}$, $\boldsymbol{\lambda}$

$$\min_{\mathbf{w}} \ \frac{1}{2} \mathbf{w}^{\mathsf{T}} \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix} \mathbf{w} + \mathbf{w}^{\mathsf{T}} \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\text{s.t. } g(\mathbf{w}) = \mathbf{w}^{\mathsf{T}} \mathbf{w} - 1 = 0$$

Guess $\lambda = 0$, step $t = 1$



Newton Iteration for Optimization - Example

Algorithm: Newton method

Input: guess $\mathbf{w}$, $\boldsymbol{\lambda}$

while $\|\nabla \mathcal{L}\|$ *or* $\|\mathbf{g}\| \geq \mathrm{tol}$ **do**

Compute

$$H(\mathbf{w}, \boldsymbol{\lambda}),\ \nabla \mathbf{g}(\mathbf{w}),\ \nabla \Phi(\mathbf{w}),\ \mathbf{g}(\mathbf{w})$$

Compute Newton direction

$$\begin{bmatrix} H & \nabla \mathbf{g} \\ \nabla \mathbf{g}^{\mathsf{T}} & 0 \end{bmatrix} \begin{bmatrix} \Delta \mathbf{w} \\ \boldsymbol{\lambda}^{+} \end{bmatrix} = - \begin{bmatrix} \nabla \Phi \\ \mathbf{g} \end{bmatrix}$$

$$\Delta \boldsymbol{\lambda} = \boldsymbol{\lambda}^{+} - \boldsymbol{\lambda}$$

Compute Newton step, $t \in]0, 1]$

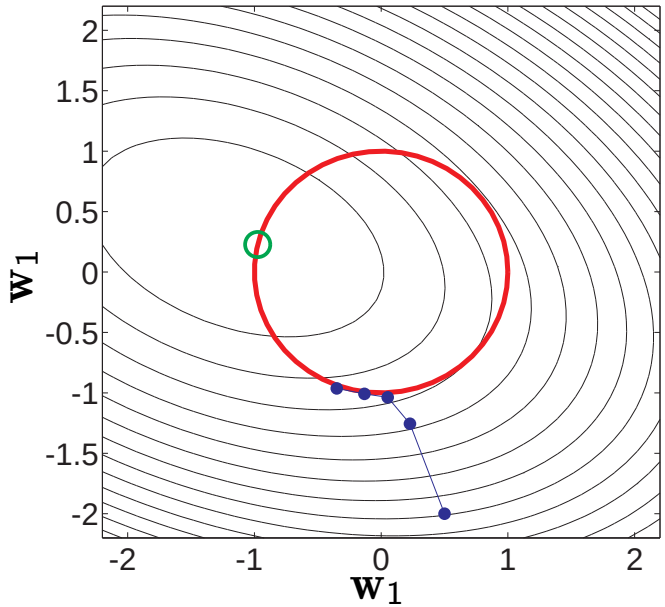
$$\mathbf{w} \leftarrow \mathbf{w} + t \Delta \mathbf{w}, \quad \boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda} + t \Delta \boldsymbol{\lambda}$$

return $\mathbf{w}$, $\boldsymbol{\lambda}$

$$\min_{\mathbf{w}} \ \frac{1}{2} \mathbf{w}^{\mathsf{T}} \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix} \mathbf{w} + \mathbf{w}^{\mathsf{T}} \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\text{s.t. } g(\mathbf{w}) = \mathbf{w}^{\mathsf{T}} \mathbf{w} - 1 = 0$$

Guess $\lambda = 0$, step $t = 1$



Newton Iteration for Optimization - Example

Algorithm: Newton method

Input: guess $\mathbf{w}$, $\boldsymbol{\lambda}$

while $\|\nabla \mathcal{L}\|$ *or* $\|\mathbf{g}\| \geq \text{tol}$ **do**

Compute

$$H(\mathbf{w}, \boldsymbol{\lambda}),\ \nabla \mathbf{g}(\mathbf{w}),\ \nabla \Phi(\mathbf{w}),\ \mathbf{g}(\mathbf{w})$$

Compute Newton direction

$$\begin{bmatrix} H & \nabla \mathbf{g} \\ \nabla \mathbf{g}^T & 0 \end{bmatrix} \begin{bmatrix} \Delta \mathbf{w} \\ \boldsymbol{\lambda}^+ \end{bmatrix} = - \begin{bmatrix} \nabla \Phi \\ \mathbf{g} \end{bmatrix}$$

$$\Delta \boldsymbol{\lambda} = \boldsymbol{\lambda}^+ - \boldsymbol{\lambda}$$

Compute Newton step, $t \in]0, 1]$

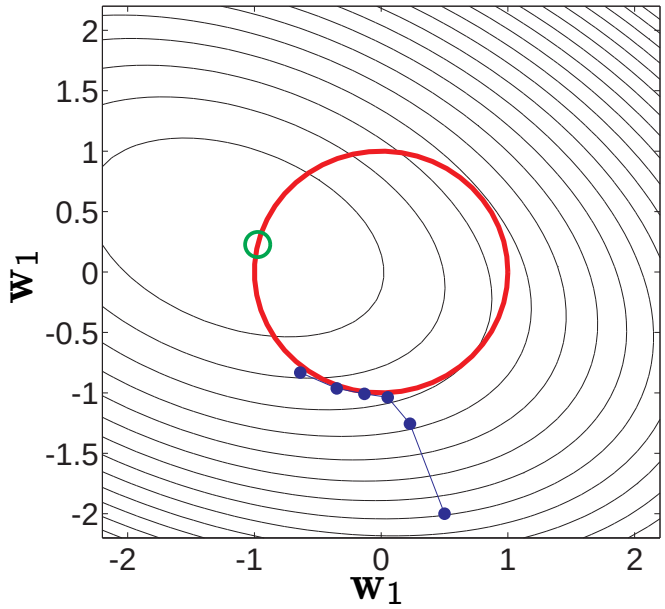
$$\mathbf{w} \leftarrow \mathbf{w} + t \Delta \mathbf{w}, \quad \boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda} + t \Delta \boldsymbol{\lambda}$$

return $\mathbf{w}$, $\boldsymbol{\lambda}$

$$\min_{\mathbf{w}} \ \frac{1}{2} \mathbf{w}^T \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix} \mathbf{w} + \mathbf{w}^T \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\text{s.t. } g(\mathbf{w}) = \mathbf{w}^T \mathbf{w} - 1 = 0$$

Guess $\lambda = 0$, step $t = 1$



Newton Iteration for Optimization - Example

Algorithm: Newton method

Input: guess $\mathbf{w}$, $\boldsymbol{\lambda}$

while $\|\nabla \mathcal{L}\|$ *or* $\|\mathbf{g}\| \geq \mathrm{tol}$ **do**

Compute

$$H(\mathbf{w}, \boldsymbol{\lambda}),\ \nabla \mathbf{g}(\mathbf{w}),\ \nabla \Phi(\mathbf{w}),\ \mathbf{g}(\mathbf{w})$$

Compute Newton direction

$$\begin{bmatrix} H & \nabla \mathbf{g} \\ \nabla \mathbf{g}^{\mathsf{T}} & 0 \end{bmatrix} \begin{bmatrix} \Delta \mathbf{w} \\ \boldsymbol{\lambda}^{+} \end{bmatrix} = - \begin{bmatrix} \nabla \Phi \\ \mathbf{g} \end{bmatrix}$$

$$\Delta \boldsymbol{\lambda} = \boldsymbol{\lambda}^{+} - \boldsymbol{\lambda}$$

Compute Newton step, $t \in]0, 1]$

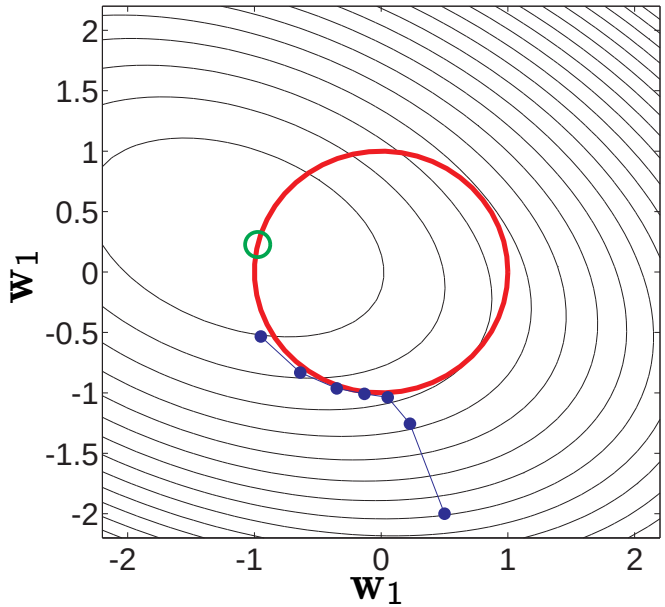
$$\mathbf{w} \leftarrow \mathbf{w} + t \Delta \mathbf{w}, \quad \boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda} + t \Delta \boldsymbol{\lambda}$$

return $\mathbf{w}$, $\boldsymbol{\lambda}$

$$\min_{\mathbf{w}} \ \frac{1}{2} \mathbf{w}^{\mathsf{T}} \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix} \mathbf{w} + \mathbf{w}^{\mathsf{T}} \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\text{s.t. } g(\mathbf{w}) = \mathbf{w}^{\mathsf{T}} \mathbf{w} - 1 = 0$$

Guess $\lambda = 0$, step $t = 1$



Newton Iteration for Optimization - Example

Algorithm: Newton method

Input: guess $\mathbf{w}$, $\boldsymbol{\lambda}$

while $\|\nabla \mathcal{L}\|$ *or* $\|\mathbf{g}\| \geq \mathrm{tol}$ **do**

Compute

$$H(\mathbf{w}, \boldsymbol{\lambda}),\ \nabla \mathbf{g}(\mathbf{w}),\ \nabla \Phi(\mathbf{w}),\ \mathbf{g}(\mathbf{w})$$

Compute Newton direction

$$\begin{bmatrix} H & \nabla \mathbf{g} \\ \nabla \mathbf{g}^{\mathsf{T}} & 0 \end{bmatrix} \begin{bmatrix} \Delta \mathbf{w} \\ \boldsymbol{\lambda}^{+} \end{bmatrix} = - \begin{bmatrix} \nabla \Phi \\ \mathbf{g} \end{bmatrix}$$

$$\Delta \boldsymbol{\lambda} = \boldsymbol{\lambda}^{+} - \boldsymbol{\lambda}$$

Compute Newton step, $t \in]0, 1]$

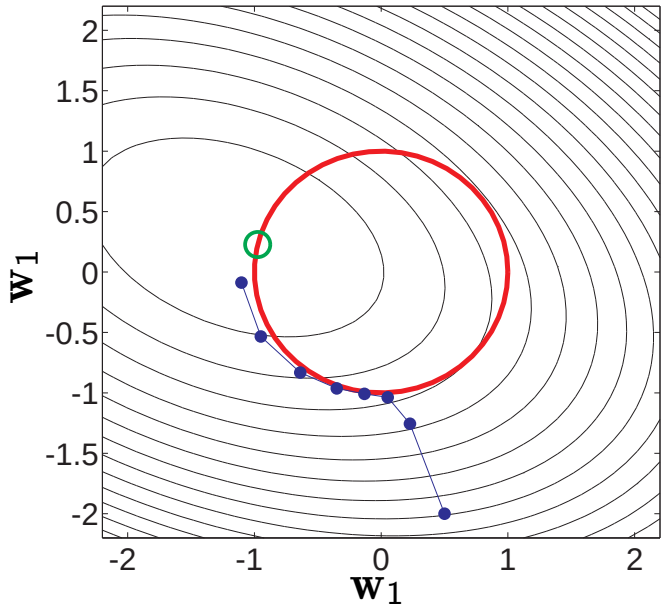
$$\mathbf{w} \leftarrow \mathbf{w} + t \Delta \mathbf{w}, \quad \boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda} + t \Delta \boldsymbol{\lambda}$$

return $\mathbf{w}$, $\boldsymbol{\lambda}$

$$\min_{\mathbf{w}} \ \frac{1}{2} \mathbf{w}^{\mathsf{T}} \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix} \mathbf{w} + \mathbf{w}^{\mathsf{T}} \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\text{s.t. } g(\mathbf{w}) = \mathbf{w}^{\mathsf{T}} \mathbf{w} - 1 = 0$$

Guess $\lambda = 0$, step $t = 1$



Newton Iteration for Optimization - Example

Algorithm: Newton method

Input: guess $\mathbf{w}$, $\boldsymbol{\lambda}$

while $\|\nabla \mathcal{L}\|$ *or* $\|\mathbf{g}\| \geq \mathrm{tol}$ **do**

Compute

$$H(\mathbf{w}, \boldsymbol{\lambda}),\ \nabla \mathbf{g}(\mathbf{w}),\ \nabla \Phi(\mathbf{w}),\ \mathbf{g}(\mathbf{w})$$

Compute Newton direction

$$\begin{bmatrix} H & \nabla \mathbf{g} \\ \nabla \mathbf{g}^{\mathsf{T}} & 0 \end{bmatrix} \begin{bmatrix} \Delta \mathbf{w} \\ \boldsymbol{\lambda}^{+} \end{bmatrix} = - \begin{bmatrix} \nabla \Phi \\ \mathbf{g} \end{bmatrix}$$

$$\Delta \boldsymbol{\lambda} = \boldsymbol{\lambda}^{+} - \boldsymbol{\lambda}$$

Compute Newton step, $t \in]0, 1]$

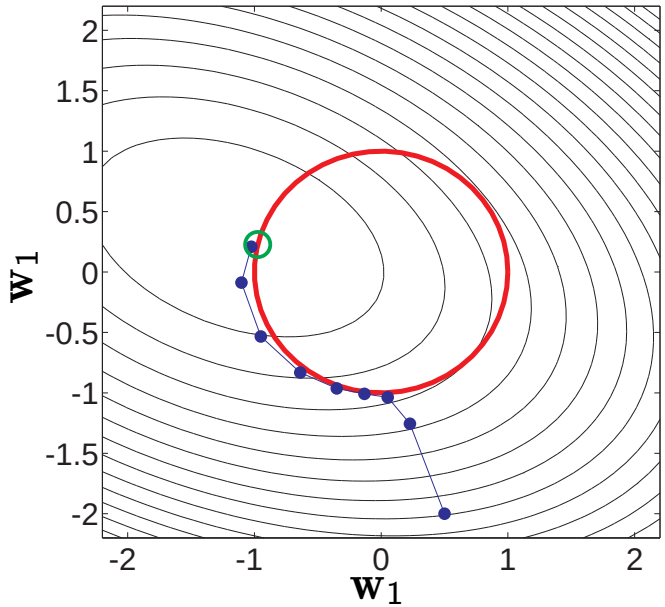
$$\mathbf{w} \leftarrow \mathbf{w} + t \Delta \mathbf{w}, \quad \boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda} + t \Delta \boldsymbol{\lambda}$$

return $\mathbf{w}$, $\boldsymbol{\lambda}$

$$\min_{\mathbf{w}} \ \frac{1}{2} \mathbf{w}^{\mathsf{T}} \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix} \mathbf{w} + \mathbf{w}^{\mathsf{T}} \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\text{s.t. } g(\mathbf{w}) = \mathbf{w}^{\mathsf{T}} \mathbf{w} - 1 = 0$$

Guess $\lambda = 0$, step $t = 1$



Newton Iteration for Optimization - Example

Algorithm: Newton method

Input: guess $\mathbf{w}$, $\boldsymbol{\lambda}$

while $\|\nabla \mathcal{L}\|$ *or* $\|\mathbf{g}\| \geq \mathrm{tol}$ **do**

Compute

$$H(\mathbf{w}, \boldsymbol{\lambda}),\ \nabla \mathbf{g}(\mathbf{w}),\ \nabla \Phi(\mathbf{w}),\ \mathbf{g}(\mathbf{w})$$

Compute Newton direction

$$\begin{bmatrix} H & \nabla \mathbf{g} \\ \nabla \mathbf{g}^{\mathsf{T}} & 0 \end{bmatrix} \begin{bmatrix} \Delta \mathbf{w} \\ \boldsymbol{\lambda}^{+} \end{bmatrix} = - \begin{bmatrix} \nabla \Phi \\ \mathbf{g} \end{bmatrix}$$

$$\Delta \boldsymbol{\lambda} = \boldsymbol{\lambda}^{+} - \boldsymbol{\lambda}$$

Compute Newton step, $t \in]0, 1]$

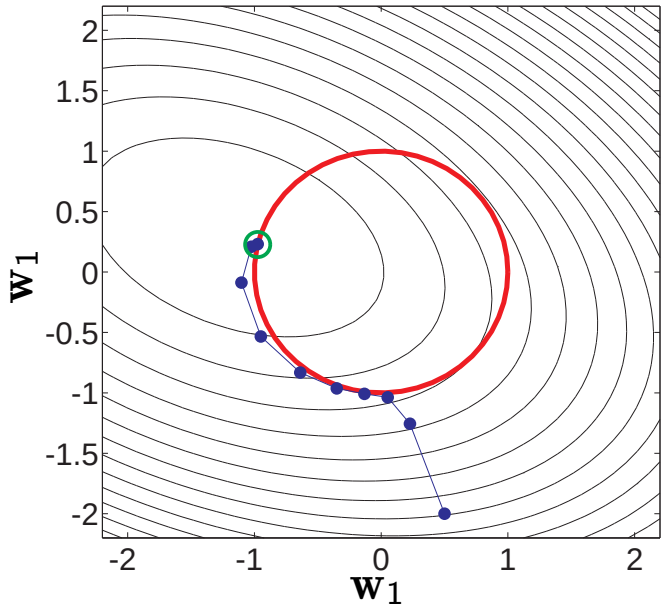
$$\mathbf{w} \leftarrow \mathbf{w} + t \Delta \mathbf{w}, \quad \boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda} + t \Delta \boldsymbol{\lambda}$$

return $\mathbf{w}$, $\boldsymbol{\lambda}$

$$\min_{\mathbf{w}} \ \frac{1}{2} \mathbf{w}^{\mathsf{T}} \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix} \mathbf{w} + \mathbf{w}^{\mathsf{T}} \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\text{s.t. } g(\mathbf{w}) = \mathbf{w}^{\mathsf{T}} \mathbf{w} - 1 = 0$$

Guess $\lambda = 0$, step $t = 1$



Newton Iteration for Optimization - Example

Algorithm: Newton method

Input: guess $\mathbf{w}$, $\boldsymbol{\lambda}$

while $\|\nabla \mathcal{L}\|$ *or* $\|\mathbf{g}\| \geq \mathrm{tol}$ **do**

Compute

$$H(\mathbf{w}, \boldsymbol{\lambda}),\ \nabla \mathbf{g}(\mathbf{w}),\ \nabla \Phi(\mathbf{w}),\ \mathbf{g}(\mathbf{w})$$

Compute Newton direction

$$\begin{bmatrix} H & \nabla \mathbf{g} \\ \nabla \mathbf{g}^{\mathsf{T}} & 0 \end{bmatrix} \begin{bmatrix} \Delta \mathbf{w} \\ \boldsymbol{\lambda}^{+} \end{bmatrix} = - \begin{bmatrix} \nabla \Phi \\ \mathbf{g} \end{bmatrix}$$

$$\Delta \boldsymbol{\lambda} = \boldsymbol{\lambda}^{+} - \boldsymbol{\lambda}$$

Compute Newton step, $t \in]0, 1]$

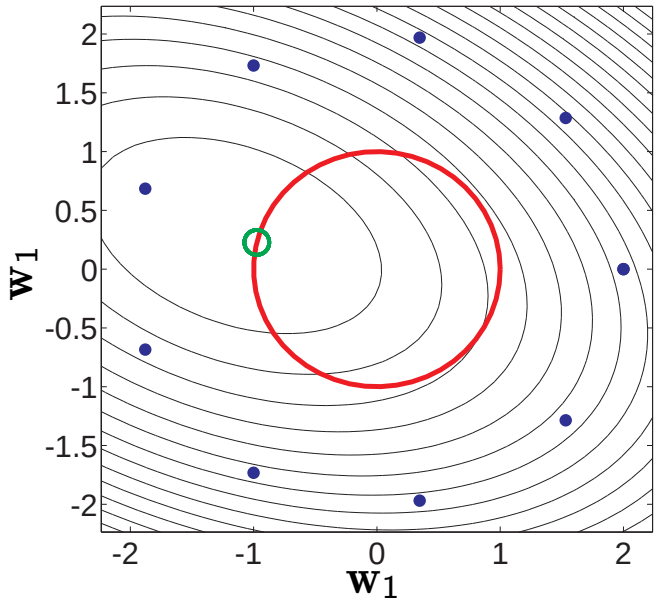
$$\mathbf{w} \leftarrow \mathbf{w} + t \Delta \mathbf{w}, \quad \boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda} + t \Delta \boldsymbol{\lambda}$$

return $\mathbf{w}$, $\boldsymbol{\lambda}$

$$\min_{\mathbf{w}} \ \frac{1}{2} \mathbf{w}^{\mathsf{T}} \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix} \mathbf{w} + \mathbf{w}^{\mathsf{T}} \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\text{s.t. } g(\mathbf{w}) = \mathbf{w}^{\mathsf{T}} \mathbf{w} - 1 = 0$$

Guess $\lambda = 0$, step $t = 1$



Newton Iteration for Optimization - Example

Algorithm: Newton method

Input: guess $\mathbf{w}$, $\boldsymbol{\lambda}$

while $\|\nabla \mathcal{L}\|$ *or* $\|\mathbf{g}\| \geq \mathrm{tol}$ **do**

Compute

$$H(\mathbf{w}, \boldsymbol{\lambda}),\ \nabla \mathbf{g}(\mathbf{w}),\ \nabla \Phi(\mathbf{w}),\ \mathbf{g}(\mathbf{w})$$

Compute Newton direction

$$\begin{bmatrix} H & \nabla \mathbf{g} \\ \nabla \mathbf{g}^{\mathsf{T}} & 0 \end{bmatrix} \begin{bmatrix} \Delta \mathbf{w} \\ \boldsymbol{\lambda}^{+} \end{bmatrix} = - \begin{bmatrix} \nabla \Phi \\ \mathbf{g} \end{bmatrix}$$

$$\Delta \boldsymbol{\lambda} = \boldsymbol{\lambda}^{+} - \boldsymbol{\lambda}$$

Compute Newton step, $t \in]0, 1]$

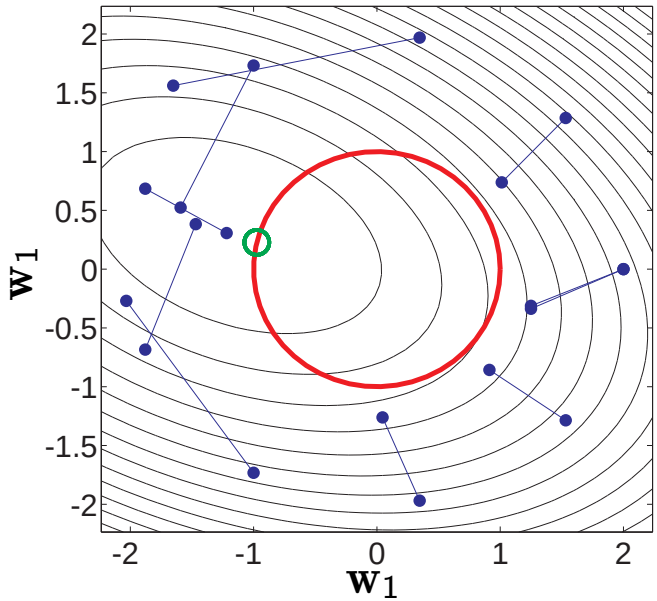
$$\mathbf{w} \leftarrow \mathbf{w} + t \Delta \mathbf{w}, \quad \boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda} + t \Delta \boldsymbol{\lambda}$$

return $\mathbf{w}$, $\boldsymbol{\lambda}$

$$\min_{\mathbf{w}} \ \frac{1}{2} \mathbf{w}^{\mathsf{T}} \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix} \mathbf{w} + \mathbf{w}^{\mathsf{T}} \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\text{s.t. } g(\mathbf{w}) = \mathbf{w}^{\mathsf{T}} \mathbf{w} - 1 = 0$$

Guess $\lambda = 0$, step $t = 1$



Newton Iteration for Optimization - Example

Algorithm: Newton method

Input: guess $\mathbf{w}$, $\boldsymbol{\lambda}$

while $\|\nabla \mathcal{L}\|$ *or* $\|\mathbf{g}\| \geq \mathrm{tol}$ **do**

Compute

$$H(\mathbf{w}, \boldsymbol{\lambda}),\ \nabla \mathbf{g}(\mathbf{w}),\ \nabla \Phi(\mathbf{w}),\ \mathbf{g}(\mathbf{w})$$

Compute Newton direction

$$\begin{bmatrix} H & \nabla \mathbf{g} \\ \nabla \mathbf{g}^{\mathsf{T}} & 0 \end{bmatrix} \begin{bmatrix} \Delta \mathbf{w} \\ \boldsymbol{\lambda}^{+} \end{bmatrix} = - \begin{bmatrix} \nabla \Phi \\ \mathbf{g} \end{bmatrix}$$

$$\Delta \boldsymbol{\lambda} = \boldsymbol{\lambda}^{+} - \boldsymbol{\lambda}$$

Compute Newton step, $t \in]0, 1]$

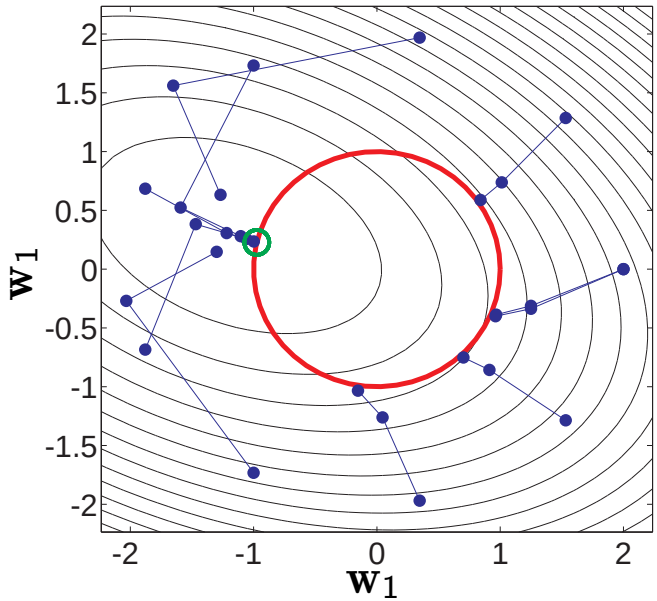
$$\mathbf{w} \leftarrow \mathbf{w} + t \Delta \mathbf{w}, \quad \boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda} + t \Delta \boldsymbol{\lambda}$$

return $\mathbf{w}$, $\boldsymbol{\lambda}$

$$\min_{\mathbf{w}} \ \frac{1}{2} \mathbf{w}^{\mathsf{T}} \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix} \mathbf{w} + \mathbf{w}^{\mathsf{T}} \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\text{s.t. } g(\mathbf{w}) = \mathbf{w}^{\mathsf{T}} \mathbf{w} - 1 = 0$$

Guess $\lambda = 0$, step $t = 1$



Newton Iteration for Optimization - Example

Algorithm: Newton method

Input: guess $\mathbf{w}$, $\boldsymbol{\lambda}$

while $\|\nabla \mathcal{L}\|$ *or* $\|\mathbf{g}\| \geq \mathrm{tol}$ **do**

Compute

$$H(\mathbf{w}, \boldsymbol{\lambda}),\ \nabla \mathbf{g}(\mathbf{w}),\ \nabla \Phi(\mathbf{w}),\ \mathbf{g}(\mathbf{w})$$

Compute Newton direction

$$\begin{bmatrix} H & \nabla \mathbf{g} \\ \nabla \mathbf{g}^{\mathsf{T}} & 0 \end{bmatrix} \begin{bmatrix} \Delta \mathbf{w} \\ \boldsymbol{\lambda}^{+} \end{bmatrix} = - \begin{bmatrix} \nabla \Phi \\ \mathbf{g} \end{bmatrix}$$

$$\Delta \boldsymbol{\lambda} = \boldsymbol{\lambda}^{+} - \boldsymbol{\lambda}$$

Compute Newton step, $t \in]0, 1]$

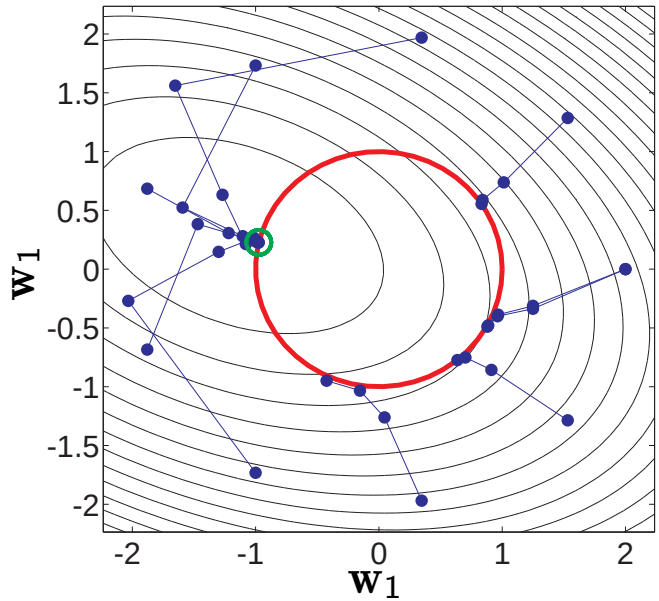
$$\mathbf{w} \leftarrow \mathbf{w} + t \Delta \mathbf{w}, \quad \boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda} + t \Delta \boldsymbol{\lambda}$$

return $\mathbf{w}$, $\boldsymbol{\lambda}$

$$\min_{\mathbf{w}} \ \frac{1}{2} \mathbf{w}^{\mathsf{T}} \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix} \mathbf{w} + \mathbf{w}^{\mathsf{T}} \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\text{s.t. } g(\mathbf{w}) = \mathbf{w}^{\mathsf{T}} \mathbf{w} - 1 = 0$$

Guess $\lambda = 0$, step $t = 1$



Newton Iteration for Optimization - Example

Algorithm: Newton method

Input: guess $\mathbf{w}$, $\boldsymbol{\lambda}$

while $\|\nabla \mathcal{L}\|$ *or* $\|\mathbf{g}\| \geq \mathrm{tol}$ **do**

Compute

$$H(\mathbf{w}, \boldsymbol{\lambda}),\ \nabla \mathbf{g}(\mathbf{w}),\ \nabla \Phi(\mathbf{w}),\ \mathbf{g}(\mathbf{w})$$

Compute Newton direction

$$\begin{bmatrix} H & \nabla \mathbf{g} \\ \nabla \mathbf{g}^T & 0 \end{bmatrix} \begin{bmatrix} \Delta \mathbf{w} \\ \boldsymbol{\lambda}^+ \end{bmatrix} = - \begin{bmatrix} \nabla \Phi \\ \mathbf{g} \end{bmatrix}$$

$$\Delta \boldsymbol{\lambda} = \boldsymbol{\lambda}^+ - \boldsymbol{\lambda}$$

Compute Newton step, $t \in]0, 1]$

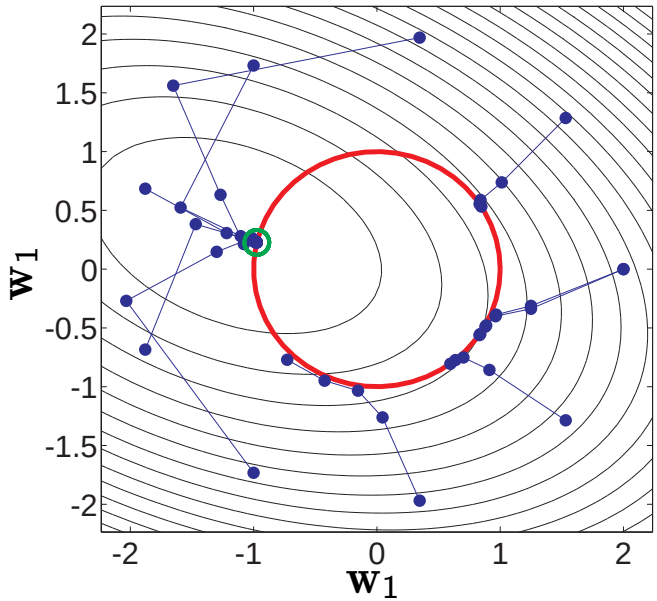
$$\mathbf{w} \leftarrow \mathbf{w} + t \Delta \mathbf{w}, \quad \boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda} + t \Delta \boldsymbol{\lambda}$$

return $\mathbf{w}$, $\boldsymbol{\lambda}$

$$\min_{\mathbf{w}} \ \frac{1}{2} \mathbf{w}^T \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix} \mathbf{w} + \mathbf{w}^T \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\text{s.t. } g(\mathbf{w}) = \mathbf{w}^T \mathbf{w} - 1 = 0$$

Guess $\lambda = 0$, step $t = 1$



Newton Iteration for Optimization - Example

Algorithm: Newton method

Input: guess $\mathbf{w}$, $\boldsymbol{\lambda}$

while $\|\nabla \mathcal{L}\|$ *or* $\|\mathbf{g}\| \geq \mathrm{tol}$ **do**

Compute

$$H(\mathbf{w}, \boldsymbol{\lambda}),\ \nabla \mathbf{g}(\mathbf{w}),\ \nabla \Phi(\mathbf{w}),\ \mathbf{g}(\mathbf{w})$$

Compute Newton direction

$$\begin{bmatrix} H & \nabla \mathbf{g} \\ \nabla \mathbf{g}^{\mathsf{T}} & 0 \end{bmatrix} \begin{bmatrix} \Delta \mathbf{w} \\ \boldsymbol{\lambda}^{+} \end{bmatrix} = - \begin{bmatrix} \nabla \Phi \\ \mathbf{g} \end{bmatrix}$$

$$\Delta \boldsymbol{\lambda} = \boldsymbol{\lambda}^{+} - \boldsymbol{\lambda}$$

Compute Newton step, $t \in]0, 1]$

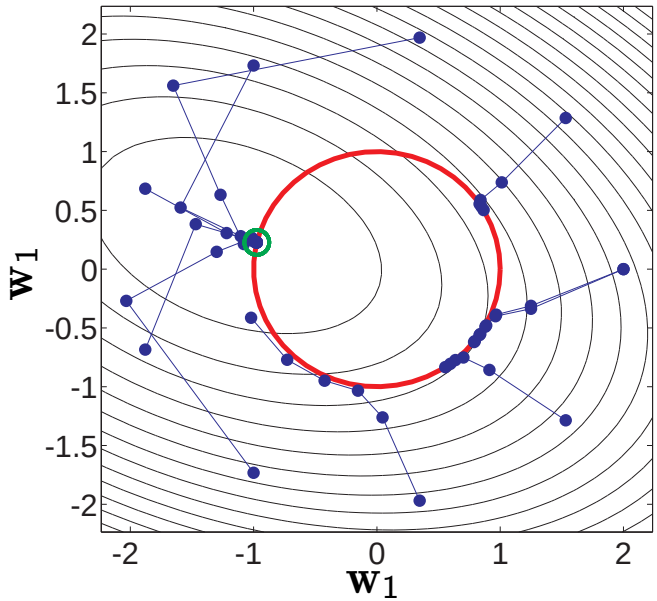
$$\mathbf{w} \leftarrow \mathbf{w} + t \Delta \mathbf{w}, \quad \boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda} + t \Delta \boldsymbol{\lambda}$$

return $\mathbf{w}$, $\boldsymbol{\lambda}$

$$\min_{\mathbf{w}} \ \frac{1}{2} \mathbf{w}^{\mathsf{T}} \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix} \mathbf{w} + \mathbf{w}^{\mathsf{T}} \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\text{s.t. } g(\mathbf{w}) = \mathbf{w}^{\mathsf{T}} \mathbf{w} - 1 = 0$$

Guess $\lambda = 0$, step $t = 1$



Newton Iteration for Optimization - Example

Algorithm: Newton method

Input: guess $\mathbf{w}$, $\boldsymbol{\lambda}$

while $\|\nabla \mathcal{L}\|$ *or* $\|\mathbf{g}\| \geq \text{tol}$ **do**

Compute

$$H(\mathbf{w}, \boldsymbol{\lambda}),\ \nabla \mathbf{g}(\mathbf{w}),\ \nabla \Phi(\mathbf{w}),\ \mathbf{g}(\mathbf{w})$$

Compute Newton direction

$$\begin{bmatrix} H & \nabla \mathbf{g} \\ \nabla \mathbf{g}^T & 0 \end{bmatrix} \begin{bmatrix} \Delta \mathbf{w} \\ \boldsymbol{\lambda}^+ \end{bmatrix} = - \begin{bmatrix} \nabla \Phi \\ \mathbf{g} \end{bmatrix}$$

$$\Delta \boldsymbol{\lambda} = \boldsymbol{\lambda}^+ - \boldsymbol{\lambda}$$

Compute Newton step, $t \in]0, 1]$

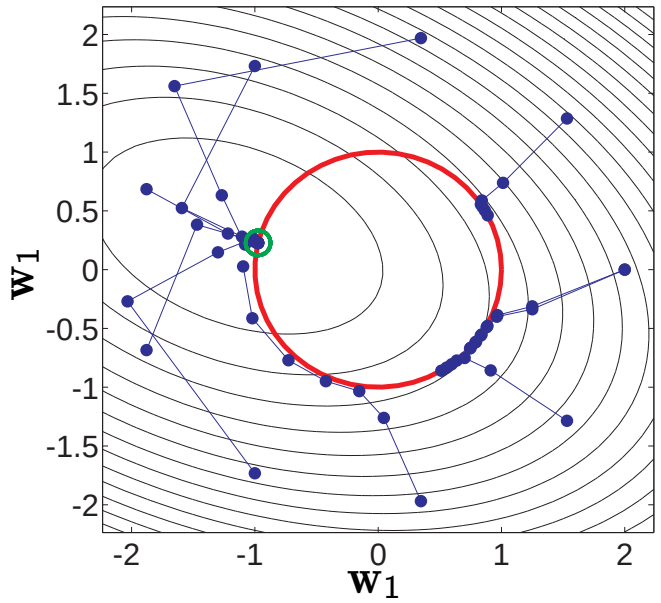
$$\mathbf{w} \leftarrow \mathbf{w} + t \Delta \mathbf{w}, \quad \boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda} + t \Delta \boldsymbol{\lambda}$$

return $\mathbf{w}$, $\boldsymbol{\lambda}$

$$\min_{\mathbf{w}} \ \frac{1}{2} \mathbf{w}^T \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix} \mathbf{w} + \mathbf{w}^T \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\text{s.t. } g(\mathbf{w}) = \mathbf{w}^T \mathbf{w} - 1 = 0$$

Guess $\lambda = 0$, step $t = 1$



Newton Iteration for Optimization - Example

Algorithm: Newton method

Input: guess $\mathbf{w}$, $\boldsymbol{\lambda}$

while $\|\nabla \mathcal{L}\|$ *or* $\|\mathbf{g}\| \geq \mathrm{tol}$ **do**

Compute

$$H(\mathbf{w}, \boldsymbol{\lambda}),\ \nabla \mathbf{g}(\mathbf{w}),\ \nabla \Phi(\mathbf{w}),\ \mathbf{g}(\mathbf{w})$$

Compute Newton direction

$$\begin{bmatrix} H & \nabla \mathbf{g} \\ \nabla \mathbf{g}^T & 0 \end{bmatrix} \begin{bmatrix} \Delta \mathbf{w} \\ \boldsymbol{\lambda}^+ \end{bmatrix} = - \begin{bmatrix} \nabla \Phi \\ \mathbf{g} \end{bmatrix}$$

$$\Delta \boldsymbol{\lambda} = \boldsymbol{\lambda}^+ - \boldsymbol{\lambda}$$

Compute Newton step, $t \in]0, 1]$

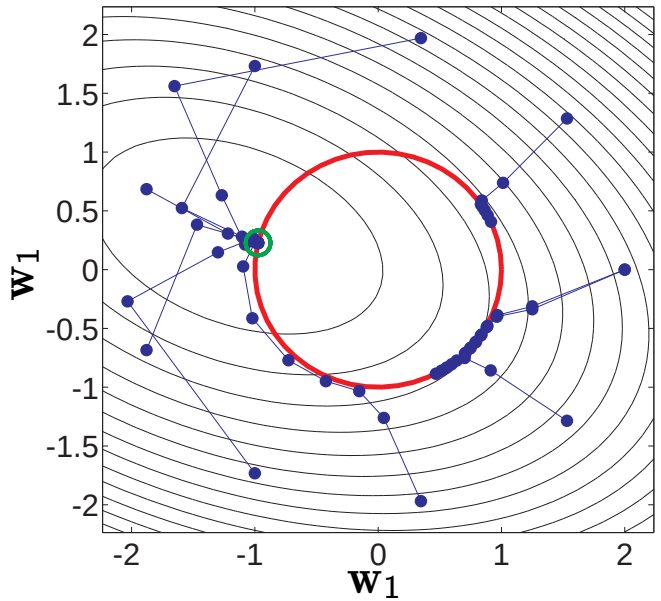
$$\mathbf{w} \leftarrow \mathbf{w} + t \Delta \mathbf{w}, \quad \boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda} + t \Delta \boldsymbol{\lambda}$$

return $\mathbf{w}$, $\boldsymbol{\lambda}$

$$\min_{\mathbf{w}} \ \frac{1}{2} \mathbf{w}^T \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix} \mathbf{w} + \mathbf{w}^T \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\text{s.t. } g(\mathbf{w}) = \mathbf{w}^T \mathbf{w} - 1 = 0$$

Guess $\lambda = 0$, step $t = 1$



Newton Iteration for Optimization - Example

Algorithm: Newton method

Input: guess $\mathbf{w}$, $\boldsymbol{\lambda}$

while $\|\nabla \mathcal{L}\|$ *or* $\|\mathbf{g}\| \geq \text{tol}$ **do**

Compute

$$H(\mathbf{w}, \boldsymbol{\lambda}),\ \nabla \mathbf{g}(\mathbf{w}),\ \nabla \Phi(\mathbf{w}),\ \mathbf{g}(\mathbf{w})$$

Compute Newton direction

$$\begin{bmatrix} H & \nabla \mathbf{g} \\ \nabla \mathbf{g}^T & 0 \end{bmatrix} \begin{bmatrix} \Delta \mathbf{w} \\ \boldsymbol{\lambda}^+ \end{bmatrix} = - \begin{bmatrix} \nabla \Phi \\ \mathbf{g} \end{bmatrix}$$

$$\Delta \boldsymbol{\lambda} = \boldsymbol{\lambda}^+ - \boldsymbol{\lambda}$$

Compute Newton step, $t \in]0, 1]$

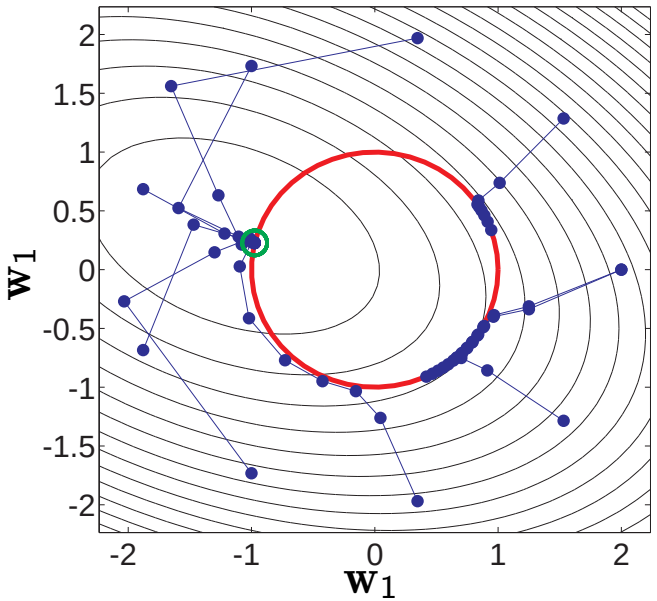
$$\mathbf{w} \leftarrow \mathbf{w} + t \Delta \mathbf{w}, \quad \boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda} + t \Delta \boldsymbol{\lambda}$$

return $\mathbf{w}$, $\boldsymbol{\lambda}$

$$\min_{\mathbf{w}} \ \frac{1}{2} \mathbf{w}^T \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix} \mathbf{w} + \mathbf{w}^T \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\text{s.t. } g(\mathbf{w}) = \mathbf{w}^T \mathbf{w} - 1 = 0$$

Guess $\lambda = 0$, step $t = 1$



Newton Iteration for Optimization - Example

Algorithm: Newton method

Input: guess $\mathbf{w}$, $\boldsymbol{\lambda}$

while $\|\nabla \mathcal{L}\|$ *or* $\|\mathbf{g}\| \geq \text{tol}$ **do**

Compute

$$H(\mathbf{w}, \boldsymbol{\lambda}),\ \nabla \mathbf{g}(\mathbf{w}),\ \nabla \Phi(\mathbf{w}),\ \mathbf{g}(\mathbf{w})$$

Compute Newton direction

$$\begin{bmatrix} H & \nabla \mathbf{g} \\ \nabla \mathbf{g}^T & 0 \end{bmatrix} \begin{bmatrix} \Delta \mathbf{w} \\ \boldsymbol{\lambda}^+ \end{bmatrix} = - \begin{bmatrix} \nabla \Phi \\ \mathbf{g} \end{bmatrix}$$

$$\Delta \boldsymbol{\lambda} = \boldsymbol{\lambda}^+ - \boldsymbol{\lambda}$$

Compute Newton step, $t \in]0, 1]$

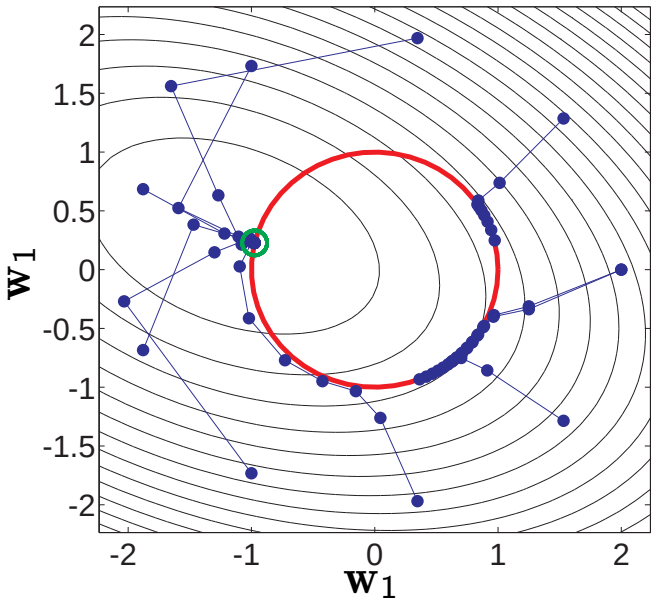
$$\mathbf{w} \leftarrow \mathbf{w} + t \Delta \mathbf{w}, \quad \boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda} + t \Delta \boldsymbol{\lambda}$$

return $\mathbf{w}$, $\boldsymbol{\lambda}$

$$\min_{\mathbf{w}} \ \frac{1}{2} \mathbf{w}^T \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix} \mathbf{w} + \mathbf{w}^T \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\text{s.t. } g(\mathbf{w}) = \mathbf{w}^T \mathbf{w} - 1 = 0$$

Guess $\lambda = 0$, step $t = 1$



Newton Iteration for Optimization - Example

Algorithm: Newton method

Input: guess $\mathbf{w}$, $\boldsymbol{\lambda}$

while $\|\nabla \mathcal{L}\|$ *or* $\|\mathbf{g}\| \geq \text{tol}$ **do**

Compute

$$H(\mathbf{w}, \boldsymbol{\lambda}),\ \nabla \mathbf{g}(\mathbf{w}),\ \nabla \Phi(\mathbf{w}),\ \mathbf{g}(\mathbf{w})$$

Compute Newton direction

$$\begin{bmatrix} H & \nabla \mathbf{g} \\ \nabla \mathbf{g}^T & 0 \end{bmatrix} \begin{bmatrix} \Delta \mathbf{w} \\ \boldsymbol{\lambda}^+ \end{bmatrix} = - \begin{bmatrix} \nabla \Phi \\ \mathbf{g} \end{bmatrix}$$

$$\Delta \boldsymbol{\lambda} = \boldsymbol{\lambda}^+ - \boldsymbol{\lambda}$$

Compute Newton step, $t \in]0, 1]$

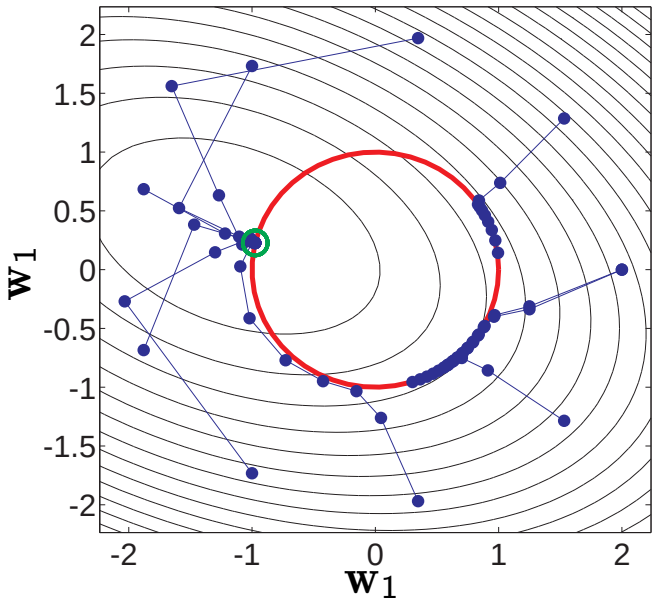
$$\mathbf{w} \leftarrow \mathbf{w} + t \Delta \mathbf{w}, \quad \boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda} + t \Delta \boldsymbol{\lambda}$$

return $\mathbf{w}$, $\boldsymbol{\lambda}$

$$\min_{\mathbf{w}} \ \frac{1}{2} \mathbf{w}^T \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix} \mathbf{w} + \mathbf{w}^T \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\text{s.t. } g(\mathbf{w}) = \mathbf{w}^T \mathbf{w} - 1 = 0$$

Guess $\lambda = 0$, step $t = 1$



Newton Iteration for Optimization - Example

Algorithm: Newton method

Input: guess $\mathbf{w}$, $\boldsymbol{\lambda}$

while $\|\nabla \mathcal{L}\|$ *or* $\|\mathbf{g}\| \geq \text{tol}$ **do**

Compute

$$H(\mathbf{w}, \boldsymbol{\lambda}),\ \nabla \mathbf{g}(\mathbf{w}),\ \nabla \Phi(\mathbf{w}),\ \mathbf{g}(\mathbf{w})$$

Compute Newton direction

$$\begin{bmatrix} H & \nabla \mathbf{g} \\ \nabla \mathbf{g}^T & 0 \end{bmatrix} \begin{bmatrix} \Delta \mathbf{w} \\ \boldsymbol{\lambda}^+ \end{bmatrix} = - \begin{bmatrix} \nabla \Phi \\ \mathbf{g} \end{bmatrix}$$

$$\Delta \boldsymbol{\lambda} = \boldsymbol{\lambda}^+ - \boldsymbol{\lambda}$$

Compute Newton step, $t \in]0, 1]$

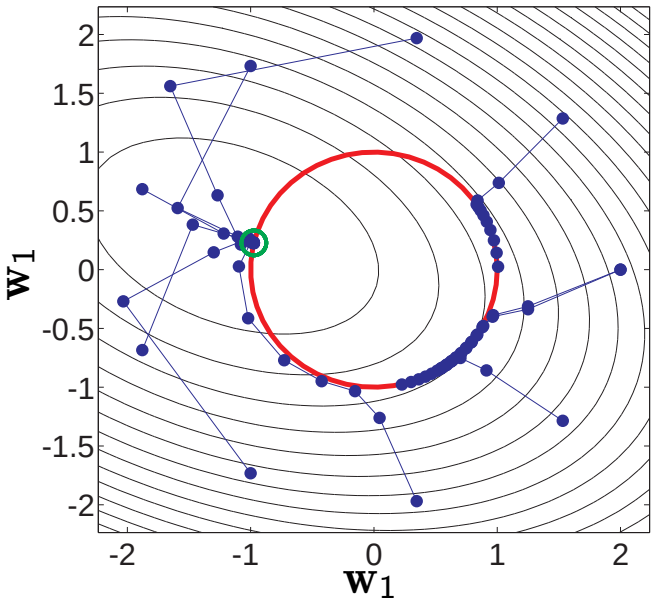
$$\mathbf{w} \leftarrow \mathbf{w} + t \Delta \mathbf{w}, \quad \boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda} + t \Delta \boldsymbol{\lambda}$$

return $\mathbf{w}$, $\boldsymbol{\lambda}$

$$\min_{\mathbf{w}} \ \frac{1}{2} \mathbf{w}^T \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix} \mathbf{w} + \mathbf{w}^T \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\text{s.t. } g(\mathbf{w}) = \mathbf{w}^T \mathbf{w} - 1 = 0$$

Guess $\lambda = 0$, step $t = 1$



Newton Iteration for Optimization - Example

Algorithm: Newton method

Input: guess $\mathbf{w}$, $\boldsymbol{\lambda}$

while $\|\nabla \mathcal{L}\|$ *or* $\|\mathbf{g}\| \geq \mathrm{tol}$ **do**

Compute

$$H(\mathbf{w}, \boldsymbol{\lambda}),\ \nabla \mathbf{g}(\mathbf{w}),\ \nabla \Phi(\mathbf{w}),\ \mathbf{g}(\mathbf{w})$$

Compute Newton direction

$$\begin{bmatrix} H & \nabla \mathbf{g} \\ \nabla \mathbf{g}^T & 0 \end{bmatrix} \begin{bmatrix} \Delta \mathbf{w} \\ \boldsymbol{\lambda}^+ \end{bmatrix} = - \begin{bmatrix} \nabla \Phi \\ \mathbf{g} \end{bmatrix}$$

$$\Delta \boldsymbol{\lambda} = \boldsymbol{\lambda}^+ - \boldsymbol{\lambda}$$

Compute Newton step, $t \in]0, 1]$

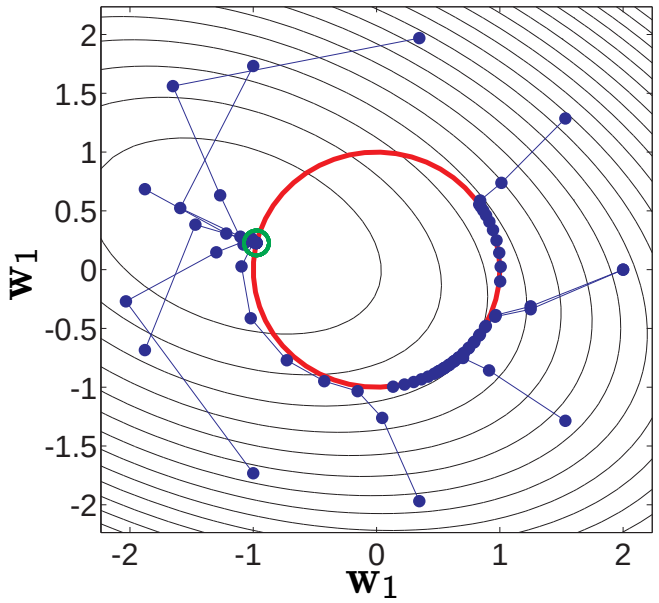
$$\mathbf{w} \leftarrow \mathbf{w} + t \Delta \mathbf{w}, \quad \boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda} + t \Delta \boldsymbol{\lambda}$$

return $\mathbf{w}$, $\boldsymbol{\lambda}$

$$\min_{\mathbf{w}} \ \frac{1}{2} \mathbf{w}^T \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix} \mathbf{w} + \mathbf{w}^T \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\text{s.t. } g(\mathbf{w}) = \mathbf{w}^T \mathbf{w} - 1 = 0$$

Guess $\lambda = 0$, step $t = 1$



Newton Iteration for Optimization - Example

Algorithm: Newton method

Input: guess $\mathbf{w}$, $\boldsymbol{\lambda}$

while $\|\nabla \mathcal{L}\|$ *or* $\|\mathbf{g}\| \geq \text{tol}$ **do**

Compute

$$H(\mathbf{w}, \boldsymbol{\lambda}),\ \nabla \mathbf{g}(\mathbf{w}),\ \nabla \Phi(\mathbf{w}),\ \mathbf{g}(\mathbf{w})$$

Compute Newton direction

$$\begin{bmatrix} H & \nabla \mathbf{g} \\ \nabla \mathbf{g}^T & 0 \end{bmatrix} \begin{bmatrix} \Delta \mathbf{w} \\ \boldsymbol{\lambda}^+ \end{bmatrix} = - \begin{bmatrix} \nabla \Phi \\ \mathbf{g} \end{bmatrix}$$

$$\Delta \boldsymbol{\lambda} = \boldsymbol{\lambda}^+ - \boldsymbol{\lambda}$$

Compute Newton step, $t \in]0, 1]$

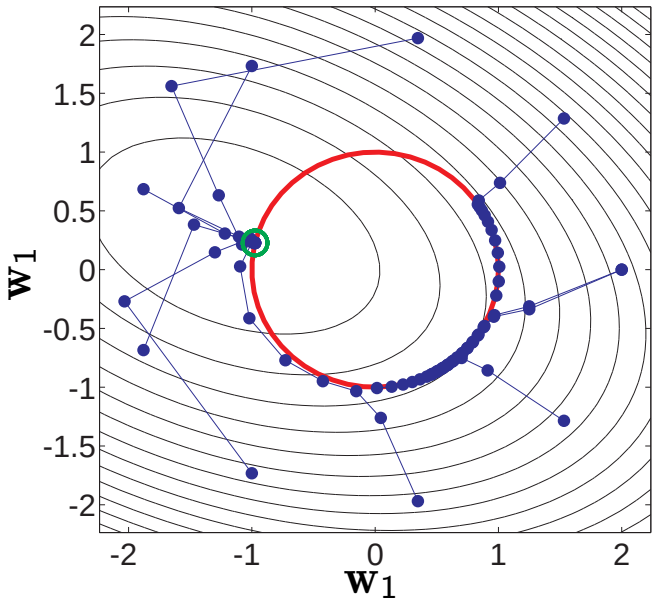
$$\mathbf{w} \leftarrow \mathbf{w} + t \Delta \mathbf{w}, \quad \boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda} + t \Delta \boldsymbol{\lambda}$$

return $\mathbf{w}$, $\boldsymbol{\lambda}$

$$\min_{\mathbf{w}} \ \frac{1}{2} \mathbf{w}^T \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix} \mathbf{w} + \mathbf{w}^T \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\text{s.t. } g(\mathbf{w}) = \mathbf{w}^T \mathbf{w} - 1 = 0$$

Guess $\lambda = 0$, step $t = 1$



Newton Iteration for Optimization - Example

Algorithm: Newton method

Input: guess $\mathbf{w}$, $\boldsymbol{\lambda}$

while $\|\nabla \mathcal{L}\|$ *or* $\|\mathbf{g}\| \geq \text{tol}$ **do**

Compute

$$H(\mathbf{w}, \boldsymbol{\lambda}),\ \nabla \mathbf{g}(\mathbf{w}),\ \nabla \Phi(\mathbf{w}),\ \mathbf{g}(\mathbf{w})$$

Compute Newton direction

$$\begin{bmatrix} H & \nabla \mathbf{g} \\ \nabla \mathbf{g}^T & 0 \end{bmatrix} \begin{bmatrix} \Delta \mathbf{w} \\ \boldsymbol{\lambda}^+ \end{bmatrix} = - \begin{bmatrix} \nabla \Phi \\ \mathbf{g} \end{bmatrix}$$

$$\Delta \boldsymbol{\lambda} = \boldsymbol{\lambda}^+ - \boldsymbol{\lambda}$$

Compute Newton step, $t \in]0, 1]$

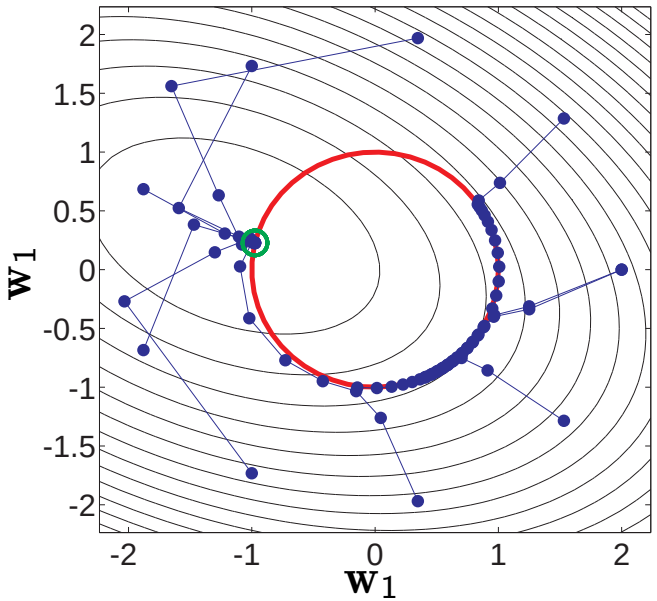
$$\mathbf{w} \leftarrow \mathbf{w} + t \Delta \mathbf{w}, \quad \boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda} + t \Delta \boldsymbol{\lambda}$$

return $\mathbf{w}$, $\boldsymbol{\lambda}$

$$\min_{\mathbf{w}} \ \frac{1}{2} \mathbf{w}^T \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix} \mathbf{w} + \mathbf{w}^T \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\text{s.t. } g(\mathbf{w}) = \mathbf{w}^T \mathbf{w} - 1 = 0$$

Guess $\lambda = 0$, step $t = 1$



Newton Iteration for Optimization - Example

Algorithm: Newton method

Input: guess $\mathbf{w}$, $\boldsymbol{\lambda}$

while $\|\nabla \mathcal{L}\|$ *or* $\|\mathbf{g}\| \geq \text{tol}$ **do**

Compute

$$H(\mathbf{w}, \boldsymbol{\lambda}),\ \nabla \mathbf{g}(\mathbf{w}),\ \nabla \Phi(\mathbf{w}),\ \mathbf{g}(\mathbf{w})$$

Compute Newton direction

$$\begin{bmatrix} H & \nabla \mathbf{g} \\ \nabla \mathbf{g}^T & 0 \end{bmatrix} \begin{bmatrix} \Delta \mathbf{w} \\ \boldsymbol{\lambda}^+ \end{bmatrix} = - \begin{bmatrix} \nabla \Phi \\ \mathbf{g} \end{bmatrix}$$

$$\Delta \boldsymbol{\lambda} = \boldsymbol{\lambda}^+ - \boldsymbol{\lambda}$$

Compute Newton step, $t \in]0, 1]$

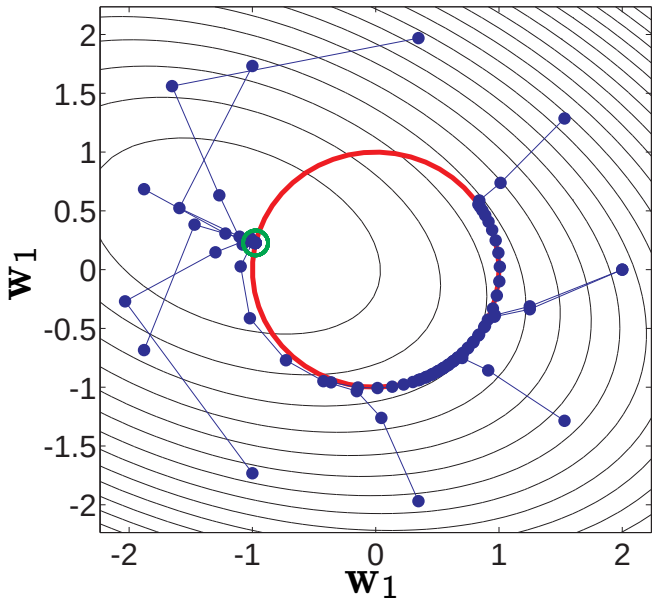
$$\mathbf{w} \leftarrow \mathbf{w} + t \Delta \mathbf{w}, \quad \boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda} + t \Delta \boldsymbol{\lambda}$$

return $\mathbf{w}$, $\boldsymbol{\lambda}$

$$\min_{\mathbf{w}} \ \frac{1}{2} \mathbf{w}^T \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix} \mathbf{w} + \mathbf{w}^T \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\text{s.t. } g(\mathbf{w}) = \mathbf{w}^T \mathbf{w} - 1 = 0$$

Guess $\lambda = 0$, step $t = 1$



Newton Iteration for Optimization - Example

Algorithm: Newton method

Input: guess $\mathbf{w}$, $\boldsymbol{\lambda}$

while $\|\nabla \mathcal{L}\|$ *or* $\|\mathbf{g}\| \geq \text{tol}$ **do**

Compute

$$H(\mathbf{w}, \boldsymbol{\lambda}),\ \nabla \mathbf{g}(\mathbf{w}),\ \nabla \Phi(\mathbf{w}),\ \mathbf{g}(\mathbf{w})$$

Compute Newton direction

$$\begin{bmatrix} H & \nabla \mathbf{g} \\ \nabla \mathbf{g}^T & 0 \end{bmatrix} \begin{bmatrix} \Delta \mathbf{w} \\ \boldsymbol{\lambda}^+ \end{bmatrix} = - \begin{bmatrix} \nabla \Phi \\ \mathbf{g} \end{bmatrix}$$

$$\Delta \boldsymbol{\lambda} = \boldsymbol{\lambda}^+ - \boldsymbol{\lambda}$$

Compute Newton step, $t \in]0, 1]$

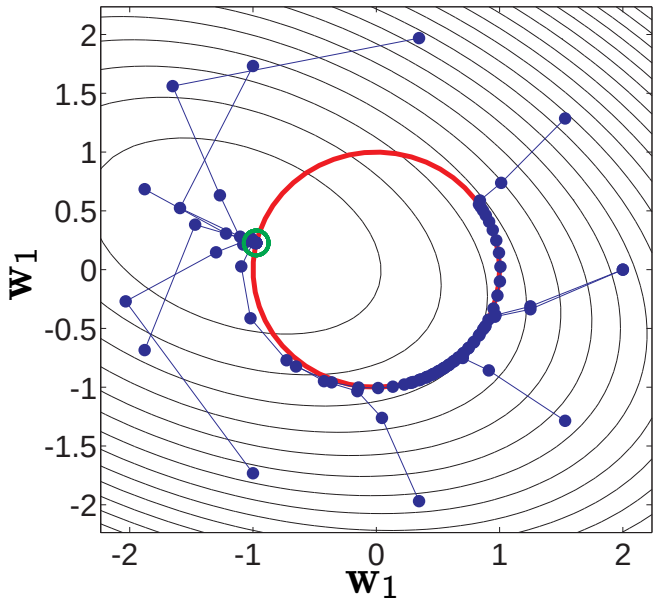
$$\mathbf{w} \leftarrow \mathbf{w} + t \Delta \mathbf{w}, \quad \boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda} + t \Delta \boldsymbol{\lambda}$$

return $\mathbf{w}$, $\boldsymbol{\lambda}$

$$\min_{\mathbf{w}} \ \frac{1}{2} \mathbf{w}^T \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix} \mathbf{w} + \mathbf{w}^T \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\text{s.t. } g(\mathbf{w}) = \mathbf{w}^T \mathbf{w} - 1 = 0$$

Guess $\lambda = 0$, step $t = 1$



Newton Iteration for Optimization - Example

Algorithm: Newton method

Input: guess $\mathbf{w}$, $\boldsymbol{\lambda}$

while $\|\nabla \mathcal{L}\|$ *or* $\|\mathbf{g}\| \geq \text{tol}$ **do**

Compute

$$H(\mathbf{w}, \boldsymbol{\lambda}),\ \nabla \mathbf{g}(\mathbf{w}),\ \nabla \Phi(\mathbf{w}),\ \mathbf{g}(\mathbf{w})$$

Compute Newton direction

$$\begin{bmatrix} H & \nabla \mathbf{g} \\ \nabla \mathbf{g}^T & 0 \end{bmatrix} \begin{bmatrix} \Delta \mathbf{w} \\ \boldsymbol{\lambda}^+ \end{bmatrix} = - \begin{bmatrix} \nabla \Phi \\ \mathbf{g} \end{bmatrix}$$

$$\Delta \boldsymbol{\lambda} = \boldsymbol{\lambda}^+ - \boldsymbol{\lambda}$$

Compute Newton step, $t \in]0, 1]$

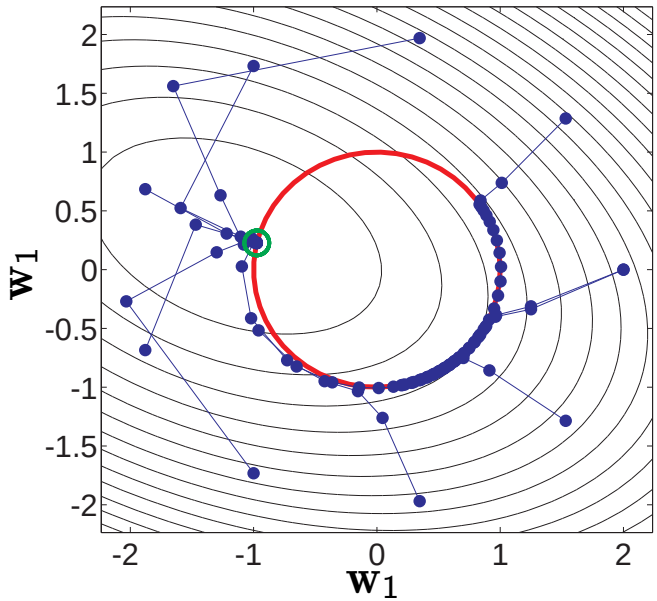
$$\mathbf{w} \leftarrow \mathbf{w} + t \Delta \mathbf{w}, \quad \boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda} + t \Delta \boldsymbol{\lambda}$$

return $\mathbf{w}$, $\boldsymbol{\lambda}$

$$\min_{\mathbf{w}} \ \frac{1}{2} \mathbf{w}^T \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix} \mathbf{w} + \mathbf{w}^T \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\text{s.t. } g(\mathbf{w}) = \mathbf{w}^T \mathbf{w} - 1 = 0$$

Guess $\lambda = 0$, step $t = 1$



Newton Iteration for Optimization - Example

Algorithm: Newton method

Input: guess $\mathbf{w}$, $\boldsymbol{\lambda}$

while $\|\nabla \mathcal{L}\|$ *or* $\|\mathbf{g}\| \geq \text{tol}$ **do**

Compute

$$H(\mathbf{w}, \boldsymbol{\lambda}),\ \nabla \mathbf{g}(\mathbf{w}),\ \nabla \Phi(\mathbf{w}),\ \mathbf{g}(\mathbf{w})$$

Compute Newton direction

$$\begin{bmatrix} H & \nabla \mathbf{g} \\ \nabla \mathbf{g}^T & 0 \end{bmatrix} \begin{bmatrix} \Delta \mathbf{w} \\ \boldsymbol{\lambda}^+ \end{bmatrix} = - \begin{bmatrix} \nabla \Phi \\ \mathbf{g} \end{bmatrix}$$

$$\Delta \boldsymbol{\lambda} = \boldsymbol{\lambda}^+ - \boldsymbol{\lambda}$$

Compute Newton step, $t \in]0, 1]$

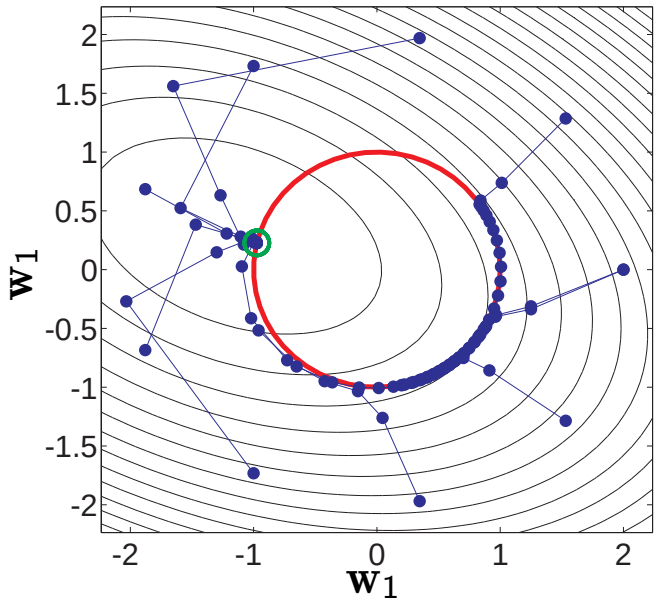
$$\mathbf{w} \leftarrow \mathbf{w} + t \Delta \mathbf{w}, \quad \boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda} + t \Delta \boldsymbol{\lambda}$$

return $\mathbf{w}$, $\boldsymbol{\lambda}$

$$\min_{\mathbf{w}} \ \frac{1}{2} \mathbf{w}^T \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix} \mathbf{w} + \mathbf{w}^T \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\text{s.t. } g(\mathbf{w}) = \mathbf{w}^T \mathbf{w} - 1 = 0$$

Guess $\lambda = 0$, step $t = 1$



Your initial guess matters !!

Invertibility of the KKT matrix

The Newton direction of the KKT conditions

$$\underbrace{\begin{bmatrix} H(\mathbf{w},\boldsymbol{\lambda}) & \nabla \mathbf{g}(\mathbf{w}) \\ \nabla \mathbf{g}(\mathbf{w})^{\mathsf{T}} & 0 \end{bmatrix}}_{\text{KKT matrix (symmetric indefinite)}} \begin{bmatrix} \Delta \mathbf{w} \\ \boldsymbol{\lambda}^{+} \end{bmatrix} = - \begin{bmatrix} \nabla \Phi(\mathbf{w}) \\ \mathbf{g}(\mathbf{w}) \end{bmatrix}$$

where $H(\mathbf{w},\boldsymbol{\lambda}) = \nabla^2_{\mathbf{w}} \mathcal{L}(\mathbf{w},\boldsymbol{\lambda})$

Invertibility of the KKT matrix

The Newton direction of the KKT conditions

$$\underbrace{\begin{bmatrix} H(\mathbf{w},\boldsymbol{\lambda}) & \nabla \mathbf{g}(\mathbf{w}) \\ \nabla \mathbf{g}(\mathbf{w})^{\mathsf{T}} & 0 \end{bmatrix}}_{\text{KKT matrix (symmetric indefinite)}} \begin{bmatrix} \Delta \mathbf{w} \\ \boldsymbol{\lambda}^{+} \end{bmatrix} = - \begin{bmatrix} \nabla \Phi(\mathbf{w}) \\ \mathbf{g}(\mathbf{w}) \end{bmatrix}$$

where $H(\mathbf{w},\boldsymbol{\lambda}) = \nabla^2_{\mathbf{w}} \mathcal{L}(\mathbf{w},\boldsymbol{\lambda})$

The KKT matrix is invertible if (sufficient, not necessary)

- $\nabla \mathbf{g}(\mathbf{w})$ is full row rank (LICQ)
- $\forall \mathbf{d} \neq 0$, such that $\nabla \mathbf{g}(\mathbf{w})^{\top} \mathbf{d} = 0$

$$\mathbf{d}^{\top} H(\mathbf{w},\boldsymbol{\lambda})\, \mathbf{d} > 0, \qquad \text{(SOSC)}$$

Invertibility of the KKT matrix

The Newton direction of the KKT conditions

$$\underbrace{\begin{bmatrix} H(\mathbf{w},\boldsymbol{\lambda}) & \nabla \mathbf{g}(\mathbf{w}) \\ \nabla \mathbf{g}(\mathbf{w})^{\mathsf{T}} & 0 \end{bmatrix}}_{\text{KKT matrix (symmetric indefinite)}} \begin{bmatrix} \Delta \mathbf{w} \\ \boldsymbol{\lambda}^{+} \end{bmatrix} = - \begin{bmatrix} \nabla \Phi(\mathbf{w}) \\ \mathbf{g}(\mathbf{w}) \end{bmatrix}$$

where $H(\mathbf{w},\boldsymbol{\lambda}) = \nabla^2_{\mathbf{w}} \mathcal{L}(\mathbf{w},\boldsymbol{\lambda})$

The KKT matrix is invertible if (sufficient, not necessary)

- $\nabla \mathbf{g}(\mathbf{w})$ is full row rank (LICQ)
- $\forall \mathbf{d} \neq 0$, such that $\nabla \mathbf{g}(\mathbf{w})^{\top} \mathbf{d} = 0$

$$\mathbf{d}^{\top} H(\mathbf{w},\boldsymbol{\lambda})\, \mathbf{d} > 0, \qquad \text{(SOSC)}$$

If $(\mathbf{w},\boldsymbol{\lambda})$ is LICQ & SOSC, then the KKT matrix is invertible in a neighborhood of $(\mathbf{w},\boldsymbol{\lambda})$

Invertibility of the KKT matrix

The Newton direction of the KKT conditions

$$\underbrace{\begin{bmatrix} H(\mathbf{w},\boldsymbol{\lambda}) & \nabla \mathbf{g}(\mathbf{w}) \\ \nabla \mathbf{g}(\mathbf{w})^{\mathsf{T}} & 0 \end{bmatrix}}_{\text{KKT matrix (symmetric indefinite)}} \begin{bmatrix} \Delta \mathbf{w} \\ \boldsymbol{\lambda}^{+} \end{bmatrix} = -\begin{bmatrix} \nabla \Phi(\mathbf{w}) \\ \mathbf{g}(\mathbf{w}) \end{bmatrix}$$

where $H(\mathbf{w},\boldsymbol{\lambda}) = \nabla^2_{\mathbf{w}} \mathcal{L}(\mathbf{w},\boldsymbol{\lambda})$

The KKT matrix is invertible if (sufficient, not necessary)

- $\nabla \mathbf{g}(\mathbf{w})$ is full row rank (LICQ)
- $\forall \mathbf{d} \neq 0$, such that $\nabla \mathbf{g}(\mathbf{w})^{\top} \mathbf{d} = 0$

$$\mathbf{d}^{\top} H(\mathbf{w},\boldsymbol{\lambda})\, \mathbf{d} > 0, \qquad \text{(SOSC)}$$

If $(\mathbf{w},\boldsymbol{\lambda})$ is LICQ & SOSC, then the KKT matrix is invertible in a neighborhood of $(\mathbf{w},\boldsymbol{\lambda})$

Punchline: if you don't have LICQ/SOSC at $\mathbf{w}^*$, then the Newton iteration becomes (in general) ill-defined as you approach $\mathbf{w}^*$

Inertia of the KKT matrix

Matrix Inertia: # of positive, zero, negative eigenvalue (n_+, n_0, n_-)

Inertia of the KKT matrix

Matrix Inertia: # of positive, zero, negative eigenvalue (n_+, n_0, n_-)

NLP: with $\mathbf{w} \in \mathbb{R}^n$ and $\mathbf{g} \in \mathbb{R}^m$

$$\begin{aligned} \min_{\mathbf{w}} \quad & \Phi(\mathbf{w}) \\ \text{s.t.} \quad & \mathbf{g}(\mathbf{w}) = 0 \end{aligned}$$

Inertia of the KKT matrix

Matrix Inertia: # of positive, zero, negative eigenvalue (n_+, n_0, n_-)

NLP: with $\mathbf{w} \in \mathbb{R}^n$ and $\mathbf{g} \in \mathbb{R}^m$

$$\begin{aligned} \min_{\mathbf{w}} \quad & \Phi(\mathbf{w}) \\ \text{s.t.} \quad & \mathbf{g}(\mathbf{w}) = 0 \end{aligned}$$

yields **KKT matrix:**

$$M = \begin{bmatrix} H(\mathbf{w}, \boldsymbol{\lambda}) & \nabla \mathbf{g}(\mathbf{w}) \\ \nabla \mathbf{g}(\mathbf{w})^{\mathsf{T}} & 0 \end{bmatrix}$$

Inertia of the KKT matrix

Matrix Inertia: # of positive, zero, negative eigenvalue (n_+, n_0, n_-)

NLP: with $\mathbf{w} \in \mathbb{R}^n$ and $\mathbf{g} \in \mathbb{R}^m$

$$\begin{aligned} \min_{\mathbf{w}} \quad & \Phi(\mathbf{w}) \\ \text{s.t.} \quad & \mathbf{g}(\mathbf{w}) = 0 \end{aligned}$$

yields **KKT matrix:**

$$M = \begin{bmatrix} H(\mathbf{w}, \boldsymbol{\lambda}) & \nabla \mathbf{g}(\mathbf{w}) \\ \nabla \mathbf{g}(\mathbf{w})^{\mathsf{T}} & 0 \end{bmatrix}$$

Note that:

- $\mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}) = \Phi(\mathbf{w}) + \boldsymbol{\lambda}^{\top} \mathbf{g}(\mathbf{w})$

Inertia of the KKT matrix

Matrix Inertia: # of positive, zero, negative eigenvalue (n_+, n_0, n_-)

NLP: with $\mathbf{w} \in \mathbb{R}^n$ and $\mathbf{g} \in \mathbb{R}^m$

$$\begin{aligned} \min_{\mathbf{w}} \quad & \Phi(\mathbf{w}) \\ \text{s.t.} \quad & \mathbf{g}(\mathbf{w}) = 0 \end{aligned}$$

yields **KKT matrix:**

$$M = \begin{bmatrix} H(\mathbf{w}, \boldsymbol{\lambda}) & \nabla \mathbf{g}(\mathbf{w}) \\ \nabla \mathbf{g}(\mathbf{w})^{\mathsf{T}} & 0 \end{bmatrix}$$

Note that:

- $\mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}) = \Phi(\mathbf{w}) + \boldsymbol{\lambda}^{\top} \mathbf{g}(\mathbf{w})$
- $\nabla_{\mathbf{w}\boldsymbol{\lambda}} \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}) = \nabla \mathbf{g}(\mathbf{w})$

Inertia of the KKT matrix

Matrix Inertia: # of positive, zero, negative eigenvalue (n_+, n_0, n_-)

NLP: with $\mathbf{w} \in \mathbb{R}^n$ and $\mathbf{g} \in \mathbb{R}^m$

$$\begin{aligned} \min_{\mathbf{w}} \quad & \Phi(\mathbf{w}) \\ \text{s.t.} \quad & \mathbf{g}(\mathbf{w}) = 0 \end{aligned}$$

yields **KKT matrix:**

$$M = \begin{bmatrix} H(\mathbf{w}, \boldsymbol{\lambda}) & \nabla \mathbf{g}(\mathbf{w}) \\ \nabla \mathbf{g}(\mathbf{w})^{\mathsf{T}} & 0 \end{bmatrix}$$

Note that:

- $\mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}) = \Phi(\mathbf{w}) + \boldsymbol{\lambda}^{\top} \mathbf{g}(\mathbf{w})$
- $\nabla_{\mathbf{w}\boldsymbol{\lambda}} \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}) = \nabla \mathbf{g}(\mathbf{w})$
- Hence

$$\nabla \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}) = \begin{bmatrix} \nabla^2_{\mathbf{ww}} \mathcal{L} & \nabla_{\mathbf{w}\boldsymbol{\lambda}} \mathcal{L} \\ \nabla_{\boldsymbol{\lambda}\mathbf{w}} \mathcal{L} & \nabla_{\boldsymbol{\lambda}\boldsymbol{\lambda}} \mathcal{L} \end{bmatrix} = M$$

Inertia of the KKT matrix

Matrix Inertia: # of positive, zero, negative eigenvalue (n_+, n_0, n_-)

NLP: with $\mathbf{w} \in \mathbb{R}^n$ and $\mathbf{g} \in \mathbb{R}^m$

$$\begin{aligned} \min_{\mathbf{w}} \quad & \Phi(\mathbf{w}) \\ \text{s.t.} \quad & \mathbf{g}(\mathbf{w}) = 0 \end{aligned}$$

Theorem: if SOSC and LICQ hold, then the inertia of M is $(n, 0, m)$

yields **KKT matrix:**

$$M = \begin{bmatrix} H(\mathbf{w}, \boldsymbol{\lambda}) & \nabla \mathbf{g}(\mathbf{w}) \\ \nabla \mathbf{g}(\mathbf{w})^{\mathsf{T}} & 0 \end{bmatrix}$$

Note that:

- $\mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}) = \Phi(\mathbf{w}) + \boldsymbol{\lambda}^{\top} \mathbf{g}(\mathbf{w})$
- $\nabla_{\mathbf{w}\boldsymbol{\lambda}} \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}) = \nabla \mathbf{g}(\mathbf{w})$
- Hence

$$\nabla \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}) = \begin{bmatrix} \nabla^2_{\mathbf{ww}} \mathcal{L} & \nabla_{\mathbf{w}\boldsymbol{\lambda}} \mathcal{L} \\ \nabla_{\boldsymbol{\lambda}\mathbf{w}} \mathcal{L} & \nabla_{\boldsymbol{\lambda}\boldsymbol{\lambda}} \mathcal{L} \end{bmatrix} = M$$

Inertia of the KKT matrix

Matrix Inertia: # of positive, zero, negative eigenvalue (n_+, n_0, n_-)

NLP: with $\mathbf{w} \in \mathbb{R}^n$ and $\mathbf{g} \in \mathbb{R}^m$

$$\begin{aligned} \min_{\mathbf{w}} \quad & \Phi(\mathbf{w}) \\ \text{s.t.} \quad & \mathbf{g}(\mathbf{w}) = 0 \end{aligned}$$

yields **KKT matrix:**

$$M = \begin{bmatrix} H(\mathbf{w}, \boldsymbol{\lambda}) & \nabla \mathbf{g}(\mathbf{w}) \\ \nabla \mathbf{g}(\mathbf{w})^\mathsf{T} & 0 \end{bmatrix}$$

Note that:

- $\mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}) = \Phi(\mathbf{w}) + \boldsymbol{\lambda}^\top \mathbf{g}(\mathbf{w})$
- $\nabla_{\mathbf{w}\boldsymbol{\lambda}} \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}) = \nabla \mathbf{g}(\mathbf{w})$
- Hence

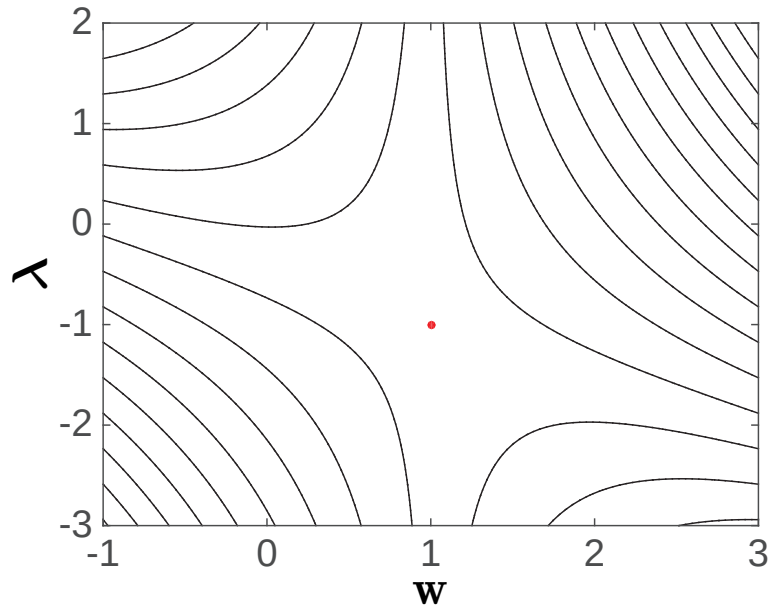
$$\nabla \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}) = \begin{bmatrix} \nabla^2_{\mathbf{ww}} \mathcal{L} & \nabla_{\mathbf{w}\boldsymbol{\lambda}} \mathcal{L} \\ \nabla_{\boldsymbol{\lambda}\mathbf{w}} \mathcal{L} & \nabla_{\boldsymbol{\lambda}\boldsymbol{\lambda}} \mathcal{L} \end{bmatrix} = M$$

Theorem: if SOSC and LICQ hold, then the inertia of M is $(n, 0, m)$

Why ?

Solution is a saddle-point of $\mathcal{L}(\mathbf{w}, \boldsymbol{\lambda})$.

- Pos. eig. match n primal directions
- Neg. eig. match m dual directions



Quadratic model interpretation

Problem:

$$\begin{array}{ll} \min\limits_{\mathbf{w}} & \Phi(\mathbf{w}) \\ \text{s.t.} & \mathbf{g}(\mathbf{w}) = 0 \end{array}$$

The Newton direction is given by

$$\begin{bmatrix} H(\mathbf{w},\boldsymbol{\lambda}) & \nabla \mathbf{g}(\mathbf{w}) \\ \nabla \mathbf{g}(\mathbf{w})^{\mathsf{T}} & 0 \end{bmatrix} \begin{bmatrix} \Delta \mathbf{w} \\ \boldsymbol{\lambda}^{+} \end{bmatrix} = - \begin{bmatrix} \nabla \Phi(\mathbf{w}) \\ \mathbf{g}(\mathbf{w}) \end{bmatrix}$$

Quadratic model interpretation

Problem:

$$\begin{array}{ll} \min\limits_{\mathbf{w}} & \Phi(\mathbf{w}) \\ \text{s.t.} & \mathbf{g}(\mathbf{w}) = 0 \end{array}$$

The Newton direction is given by

$$\begin{bmatrix} H(\mathbf{w},\boldsymbol{\lambda}) & \nabla \mathbf{g}(\mathbf{w}) \\ \nabla \mathbf{g}(\mathbf{w})^{\mathsf{T}} & 0 \end{bmatrix} \begin{bmatrix} \Delta \mathbf{w} \\ \boldsymbol{\lambda}^{+} \end{bmatrix} = - \begin{bmatrix} \nabla \Phi(\mathbf{w}) \\ \mathbf{g}(\mathbf{w}) \end{bmatrix}$$

The Newton direction is also given by the Quadratic Program (QP):

$$\begin{array}{ll} \min\limits_{\Delta \mathbf{w}} & \frac{1}{2} \Delta \mathbf{w}^{\mathsf{T}} H(\mathbf{w},\boldsymbol{\lambda}) \Delta \mathbf{w} + \nabla \Phi(\mathbf{w})^{\mathsf{T}} \Delta \mathbf{w} \\ \text{s.t.} & \mathbf{g}(\mathbf{w}) + \nabla \mathbf{g}(\mathbf{w})^{\mathsf{T}} \Delta \mathbf{w} = 0 \end{array}$$

Quadratic model interpretation

Problem:

$$\begin{array}{ll} \min\limits_{\mathbf{w}} & \Phi(\mathbf{w}) \\ \text{s.t.} & \mathbf{g}(\mathbf{w}) = 0 \end{array}$$

The Newton direction is given by

$$\begin{bmatrix} H(\mathbf{w},\boldsymbol{\lambda}) & \nabla \mathbf{g}(\mathbf{w}) \\ \nabla \mathbf{g}(\mathbf{w})^T & 0 \end{bmatrix} \begin{bmatrix} \Delta \mathbf{w} \\ \boldsymbol{\lambda}^+ \end{bmatrix} = - \begin{bmatrix} \nabla \Phi(\mathbf{w}) \\ \mathbf{g}(\mathbf{w}) \end{bmatrix}$$

The Newton direction is also given by the Quadratic Program (QP):

$$\begin{array}{ll} \min\limits_{\Delta \mathbf{w}} & \frac{1}{2} \Delta \mathbf{w}^T H(\mathbf{w},\boldsymbol{\lambda}) \Delta \mathbf{w} + \nabla \Phi(\mathbf{w})^T \Delta \mathbf{w} \\ \text{s.t.} & \mathbf{g}(\mathbf{w}) + \nabla \mathbf{g}(\mathbf{w})^T \Delta \mathbf{w} = 0 \end{array}$$

Dual variables $\boldsymbol{\lambda}^+$ given by the dual variables of the QP, i.e. $\boldsymbol{\lambda}^+ = \boldsymbol{\lambda}_{\text{QP}}$

Quadratic model interpretation

Problem:

$$\begin{array}{ll} \min\limits_{\mathbf{w}} & \Phi(\mathbf{w}) \\ \text{s.t.} & \mathbf{g}(\mathbf{w}) = 0 \end{array}$$

The Newton direction is given by

$$\begin{bmatrix} H(\mathbf{w},\boldsymbol{\lambda}) & \nabla \mathbf{g}(\mathbf{w}) \\ \nabla \mathbf{g}(\mathbf{w})^T & 0 \end{bmatrix} \begin{bmatrix} \Delta \mathbf{w} \\ \boldsymbol{\lambda}^+ \end{bmatrix} = - \begin{bmatrix} \nabla \Phi(\mathbf{w}) \\ \mathbf{g}(\mathbf{w}) \end{bmatrix}$$

The Newton direction is also given by the Quadratic Program (QP):

$$\begin{array}{ll} \min\limits_{\Delta \mathbf{w}} & \frac{1}{2} \Delta \mathbf{w}^T H(\mathbf{w},\boldsymbol{\lambda}) \Delta \mathbf{w} + \nabla \Phi(\mathbf{w})^T \Delta \mathbf{w} \\ \text{s.t.} & \mathbf{g}(\mathbf{w}) + \nabla \mathbf{g}(\mathbf{w})^T \Delta \mathbf{w} = 0 \end{array}$$

Dual variables $\boldsymbol{\lambda}^+$ given by the dual variables of the QP, i.e. $\boldsymbol{\lambda}^+ = \boldsymbol{\lambda}_{\text{QP}}$

Proof: KKT of the QP are equivalent to the KKT system.

Quadratic model interpretation

Problem:

$$\begin{array}{ll} \min\limits_{\mathbf{w}} & \Phi(\mathbf{w}) \\ \text{s.t.} & \mathbf{g}(\mathbf{w}) = 0 \end{array}$$

The Newton direction is given by

$$\begin{bmatrix} H(\mathbf{w},\boldsymbol{\lambda}) & \nabla \mathbf{g}(\mathbf{w}) \\ \nabla \mathbf{g}(\mathbf{w})^{\mathsf{T}} & 0 \end{bmatrix} \begin{bmatrix} \Delta \mathbf{w} \\ \boldsymbol{\lambda}^{+} \end{bmatrix} = - \begin{bmatrix} \nabla \Phi(\mathbf{w}) \\ \mathbf{g}(\mathbf{w}) \end{bmatrix}$$

The Newton direction is also given by the Quadratic Program (QP):

$$\begin{array}{ll} \min\limits_{\Delta \mathbf{w}} & \frac{1}{2} \Delta \mathbf{w}^{\mathsf{T}} H(\mathbf{w},\boldsymbol{\lambda}) \Delta \mathbf{w} + \nabla \Phi(\mathbf{w})^{\mathsf{T}} \Delta \mathbf{w} \\ \text{s.t.} & \mathbf{g}(\mathbf{w}) + \nabla \mathbf{g}(\mathbf{w})^{\mathsf{T}} \Delta \mathbf{w} = 0 \end{array}$$

Dual variables $\boldsymbol{\lambda}^{+}$ given by the dual variables of the QP, i.e. $\boldsymbol{\lambda}^{+} = \boldsymbol{\lambda}_{\mathrm{QP}}$

Proof: KKT of the QP are equivalent to the KKT system.

The Newton direction is given by solving a quadratic models of the original problem !!

Outline

1. The Newton method
2. Newton on the KKT conditions
3. The reduced Newton step (unconstrained problems)
4. The merit function - Line-search for constrained problems
5. Newton-type methods
6. Sequential Quadratic Programming

The Newton direction is a descent direction

Consider the unconstrained problem

$$\min_{\mathbf{w}} \quad \Phi(\mathbf{w})$$

The Newton direction $\Delta\mathbf{w}$ minimizes the quadratic model (unconstrained):

$$Q(\mathbf{w}, \Delta\mathbf{w}) = \Phi(\mathbf{w}) + \nabla\Phi(\mathbf{w})^{\mathsf{T}}\Delta\mathbf{w} + \frac{1}{2}\Delta\mathbf{w}^{\mathsf{T}} H(\mathbf{w})\Delta\mathbf{w}$$

The Newton direction is a descent direction

Consider the unconstrained problem

$$\min_{\mathbf{w}} \quad \Phi(\mathbf{w})$$

The Newton direction $\Delta\mathbf{w}$ minimizes the quadratic model (unconstrained):

$$Q(\mathbf{w}, \Delta\mathbf{w}) = \Phi(\mathbf{w}) + \nabla\Phi(\mathbf{w})^{\mathsf{T}}\Delta\mathbf{w} + \frac{1}{2}\Delta\mathbf{w}^{\mathsf{T}} H(\mathbf{w})\Delta\mathbf{w}$$

where $H(\mathbf{w}) = \nabla^2_{\mathbf{w}}\Phi(\mathbf{w})$. Step $\Delta\mathbf{w}$ reads as:

$$\Delta\mathbf{w} = -H(\mathbf{w})^{-1}\nabla\Phi(\mathbf{w})$$

The Newton direction is a descent direction

Consider the unconstrained problem

$$\min_{\mathbf{w}} \quad \Phi(\mathbf{w})$$

The Newton direction $\Delta\mathbf{w}$ minimizes the quadratic model (unconstrained):

$$Q(\mathbf{w}, \Delta\mathbf{w}) = \Phi(\mathbf{w}) + \nabla\Phi(\mathbf{w})^{\mathsf{T}}\Delta\mathbf{w} + \frac{1}{2}\Delta\mathbf{w}^{\mathsf{T}} H(\mathbf{w})\Delta\mathbf{w}$$

where $H(\mathbf{w}) = \nabla^2_{\mathbf{w}}\Phi(\mathbf{w})$. Step $\Delta\mathbf{w}$ reads as:

$$\Delta\mathbf{w} = -H(\mathbf{w})^{-1}\nabla\Phi(\mathbf{w})$$

If the Hessian $H(\mathbf{w}) \succ 0$ then for $t \to 0_+$:

$$\Phi(\mathbf{w} + t\Delta\mathbf{w}) - \Phi(\mathbf{w}) = t\nabla\Phi(\mathbf{w})^{\mathsf{T}}\Delta\mathbf{w} = -t\nabla\Phi(\mathbf{w})^{\mathsf{T}} H(\mathbf{w})^{-1}\nabla\Phi(\mathbf{w}) < 0$$

The Newton direction is a descent direction

Consider the unconstrained problem

$$\min_{\mathbf{w}} \quad \Phi(\mathbf{w})$$

The Newton direction $\Delta\mathbf{w}$ minimizes the quadratic model (unconstrained):

$$Q(\mathbf{w}, \Delta\mathbf{w}) = \Phi(\mathbf{w}) + \nabla\Phi(\mathbf{w})^T\Delta\mathbf{w} + \frac{1}{2}\Delta\mathbf{w}^T H(\mathbf{w})\Delta\mathbf{w}$$

where $H(\mathbf{w}) = \nabla^2_{\mathbf{w}}\Phi(\mathbf{w})$. Step $\Delta\mathbf{w}$ reads as:

$$\Delta\mathbf{w} = -H(\mathbf{w})^{-1}\nabla\Phi(\mathbf{w})$$

If the Hessian $H(\mathbf{w}) \succ 0$ then for $t \to 0_+$:

$$\Phi(\mathbf{w} + t\Delta\mathbf{w}) - \Phi(\mathbf{w}) = t\nabla\Phi(\mathbf{w})^T\Delta\mathbf{w} = -t\nabla\Phi(\mathbf{w})^T H(\mathbf{w})^{-1}\nabla\Phi(\mathbf{w}) < 0$$

i.e. $\boxed{\Phi(\mathbf{w} + t\Delta\mathbf{w}) < \Phi(\mathbf{w}) \text{ for } t \in \mathbb{R}_+ \text{ sufficiently small.}}$

The Newton direction is a descent direction

Consider the unconstrained problem

$$\min_{\mathbf{w}} \quad \Phi(\mathbf{w})$$

The Newton direction $\Delta\mathbf{w}$ minimizes the quadratic model (unconstrained):

$$Q(\mathbf{w}, \Delta\mathbf{w}) = \Phi(\mathbf{w}) + \nabla\Phi(\mathbf{w})^T\Delta\mathbf{w} + \frac{1}{2}\Delta\mathbf{w}^T H(\mathbf{w})\Delta\mathbf{w}$$

where $H(\mathbf{w}) = \nabla^2_{\mathbf{w}}\Phi(\mathbf{w})$. Step $\Delta\mathbf{w}$ reads as:

$$\Delta\mathbf{w} = -H(\mathbf{w})^{-1}\nabla\Phi(\mathbf{w})$$

If the Hessian $H(\mathbf{w}) \succ 0$ then for $t \to 0_+$:

$$\Phi(\mathbf{w} + t\Delta\mathbf{w}) - \Phi(\mathbf{w}) = t\nabla\Phi(\mathbf{w})^T\Delta\mathbf{w} = -t\nabla\Phi(\mathbf{w})^T H(\mathbf{w})^{-1}\nabla\Phi(\mathbf{w}) < 0$$

i.e. $\boxed{\Phi(\mathbf{w} + t\Delta\mathbf{w}) < \Phi(\mathbf{w}) \text{ for } t \in \mathbb{R}_+ \text{ sufficiently small.}}$

However, there's no guarantee that $\Phi(\mathbf{w} + \Delta\mathbf{w}) < \Phi(\mathbf{w})$!!

Failure of the full Newton step

Newton direction $\Delta \mathbf{w}$ minimizes the quadratic model :

$$Q(\mathbf{w}, \Delta \mathbf{w}) = \Phi(\mathbf{w}) + \nabla \Phi(\mathbf{w})^T \Delta \mathbf{w} + \frac{1}{2} \Delta \mathbf{w}^T H(\mathbf{w}) \Delta \mathbf{w}$$

Given by:

$$\Delta \mathbf{w} = -H(\mathbf{w})^{-1} \nabla \Phi(\mathbf{w})$$

Full Newton step:

$$\mathbf{w} \leftarrow \mathbf{w} + \Delta \mathbf{w}$$

Failure of the full Newton step

Newton direction $\Delta \mathbf{w}$ minimizes the quadratic model :

$$Q(\mathbf{w}, \Delta \mathbf{w}) = \Phi(\mathbf{w}) + \nabla \Phi(\mathbf{w})^T \Delta \mathbf{w} + \frac{1}{2} \Delta \mathbf{w}^T H(\mathbf{w}) \Delta \mathbf{w}$$

Given by:

$$\Delta \mathbf{w} = -H(\mathbf{w})^{-1} \nabla \Phi(\mathbf{w})$$

Full Newton step:

$$\mathbf{w} \leftarrow \mathbf{w} + \Delta \mathbf{w}$$

What if the quadratic model is "not good" ?

Failure of the full Newton step

Newton direction $\Delta \mathbf{w}$ minimizes the quadratic model :

$$Q(\mathbf{w}, \Delta \mathbf{w}) = \Phi(\mathbf{w}) + \nabla \Phi(\mathbf{w})^T \Delta \mathbf{w} + \frac{1}{2} \Delta \mathbf{w}^T H(\mathbf{w}) \Delta \mathbf{w}$$

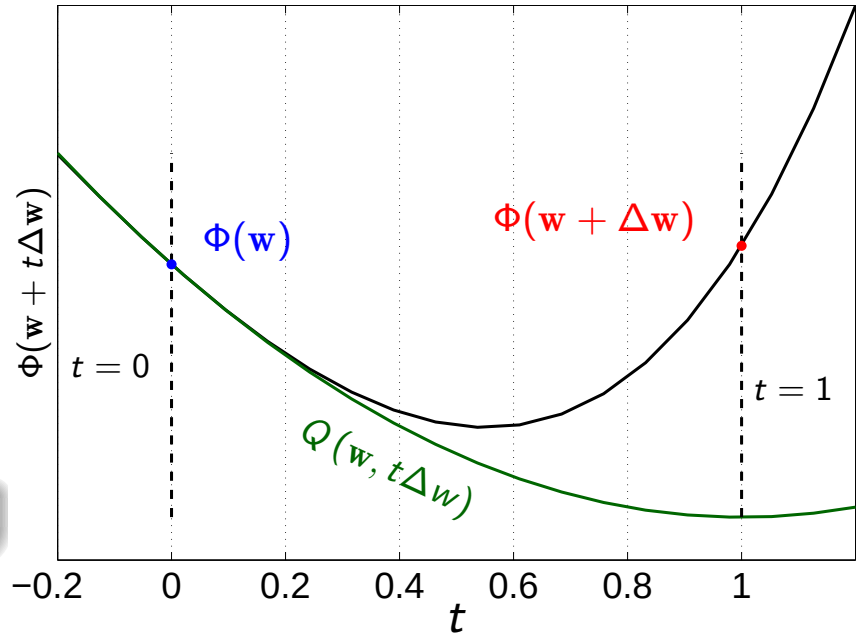
Given by:

$$\Delta \mathbf{w} = -H(\mathbf{w})^{-1} \nabla \Phi(\mathbf{w})$$

Full Newton step:

$$\mathbf{w} \leftarrow \mathbf{w} + \Delta \mathbf{w}$$

What if the quadratic model is "not good" ?



Failure of the full Newton step

Newton direction $\Delta \mathbf{w}$ minimizes the quadratic model :

$$Q(\mathbf{w}, \Delta \mathbf{w}) = \Phi(\mathbf{w}) + \nabla \Phi(\mathbf{w})^T \Delta \mathbf{w} + \frac{1}{2} \Delta \mathbf{w}^T H(\mathbf{w}) \Delta \mathbf{w}$$

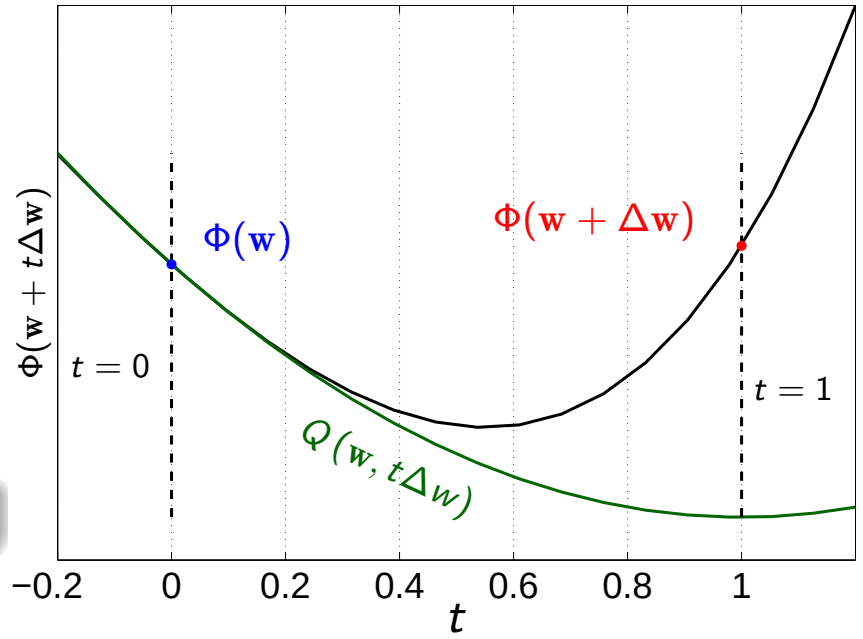
Given by:

$$\Delta \mathbf{w} = -H(\mathbf{w})^{-1} \nabla \Phi(\mathbf{w})$$

Full Newton step:

$$\mathbf{w} \leftarrow \mathbf{w} + \Delta \mathbf{w}$$

What if the quadratic model is "not good" ?



A situation with $\Phi(\mathbf{w} + \Delta \mathbf{w}) > \Phi(\mathbf{w})$ can easily occur...

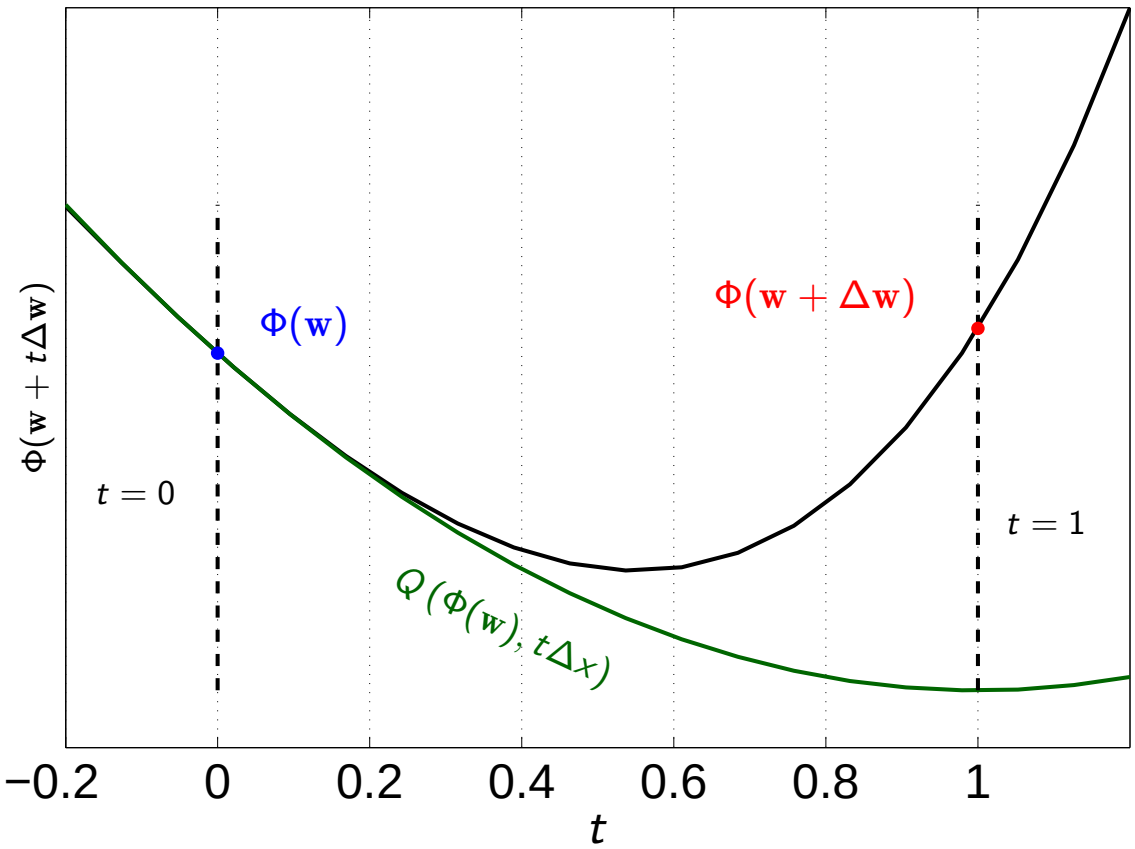
- Strong variation of H
- Nonlinear constraints (if the problem is constrained)

Globalization - Line search strategies

Exact line search (for unconstrained optimization)

Find the best step length:

$$t = \arg \min_{t \in]0,1]} \Phi(\mathbf{w} + t\Delta\mathbf{w})$$



Recap of last lecture

- Newton is a tool to solve nonlinear equations, will be used to solve the KKTs associated to an NLP
- Reduced-step Newton converges to a solution if $\nabla \mathbf{r}$ is full rank
- Convergence of full steps Newton is quadratic (fast), better when equations are "closer to linear"
- Rescaling the decision variables $\mathbf{w}$ has no effect on Newton, may help the linear algebra though
- Newton on KKT conditions has a specific structure, step well posed if LICQ & SOSC
- Newton step is equivalent to solving a QP forming a "linear-quadratic" approximation of the NLP
- In unconstrained optimization problems, this observation is equivalent to "minimizing a quadratic model of the cost", reduced steps are needed when that model validity is "locally limited"

Globalization - Line search strategies

"Armijo's" backtracking line search (for unconstrained optimization)

Given a primal direction $\Delta \mathbf{w}$, using $0 < \alpha \leq \frac{1}{2}$ and $0 < \beta < 1$, do $t = 1$:

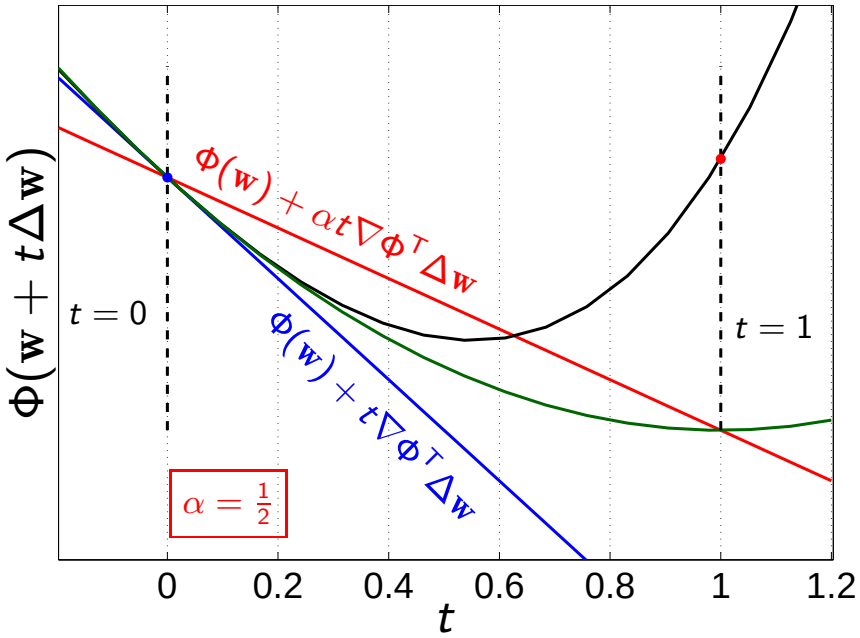
Algorithm: "Armijo's" backtracking

$t \leftarrow 1$

while $\Phi(\mathbf{w} + t\Delta\mathbf{w}) \geq \Phi(\mathbf{w}) + \alpha t \nabla \Phi(\mathbf{w})^T \Delta \mathbf{w}$ **do**

  $t \leftarrow \beta t$

return t



Globalization - Line search strategies

"Armijo's" backtracking line search (for unconstrained optimization)

Given a primal direction $\Delta \mathbf{w}$, using $0 < \alpha \leq \frac{1}{2}$ and $0 < \beta < 1$, do $t = 1$:

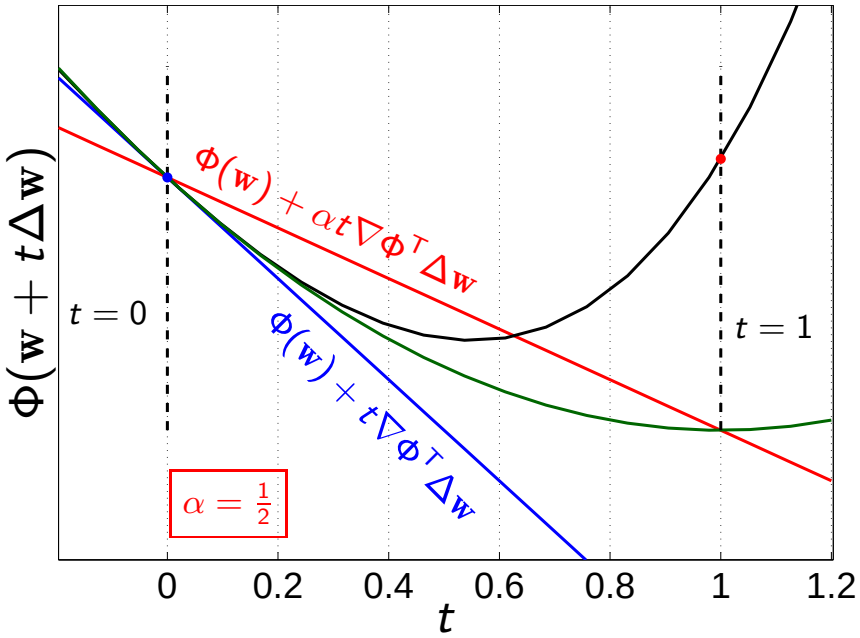
Algorithm: "Armijo's" backtracking

$t \leftarrow 1$

while $\Phi(\mathbf{w} + t\Delta\mathbf{w}) \geq \Phi(\mathbf{w}) + \alpha t \nabla\Phi(\mathbf{w})^T \Delta\mathbf{w}$ **do**

  $t \leftarrow \beta t$

return t



- If α too small we may accept steps yielding only mediocre improvement.
- If Φ is quadratic, we want full step, i.e. $\alpha \leq \frac{1}{2}$
- In practice $\alpha < \frac{1}{2}$ is usually preferred
- Additional conditions are sometimes used to yield a sufficiently large the step (c.f. Wolfe or Goldstein conditions)

Convergence of the Newton method with Line-search (unconstrained)

Theorem

Assume:

- Hessian is bounded $m \cdot I \leq \nabla^2 \Phi(\mathbf{w}) \leq M \cdot I$, with $m, M > 0$
- Hessian is Lipschitz, i.e. $\|\nabla^2 \Phi(\mathbf{w}) - \nabla^2 \Phi(\mathbf{y})\| \leq L\|\mathbf{w} - \mathbf{y}\|, \quad \forall \mathbf{w}, \mathbf{y}$

Convergence of the Newton method with Line-search (unconstrained)

Theorem

Assume:

- Hessian is bounded $m \cdot I \leq \nabla^2 \Phi(\mathbf{w}) \leq M \cdot I$, with $m, M > 0$
- Hessian is Lipschitz, i.e. $\|\nabla^2 \Phi(\mathbf{w}) - \nabla^2 \Phi(\mathbf{y})\| \leq L\|\mathbf{w} - \mathbf{y}\|, \quad \forall \mathbf{w}, \mathbf{y}$

then $\exists\ \eta, \gamma > 0$ with $\eta < \frac{m^2}{L}$ such that $\forall\ \mathbf{w}$:

Damped phase:				
$\Phi(\mathbf{w}^+) - \Phi(\mathbf{w})$	$\leq$	$-\gamma$	if	$\|\nabla \Phi(\mathbf{w})\| \geq \eta$
Quadratic phase:				
$\|\nabla \Phi(\mathbf{w}^+)\|$	$\leq$	$\frac{L}{2m^2}\|\nabla \Phi(\mathbf{w})\|^2$	if	$\|\nabla \Phi(\mathbf{w})\| < \eta$

where $\mathbf{w}^+ = \mathbf{w} + t\Delta\mathbf{w}$ is the (possibly) reduced Newton step

Convergence of the Newton method with Line-search (unconstrained)

Theorem

Assume:

- Hessian is bounded $m \cdot I \leq \nabla^2 \Phi(\mathbf{w}) \leq M \cdot I$, with $m, M > 0$
- Hessian is Lipschitz, i.e. $\|\nabla^2 \Phi(\mathbf{w}) - \nabla^2 \Phi(\mathbf{y})\| \leq L\|\mathbf{w} - \mathbf{y}\|, \quad \forall \mathbf{w}, \mathbf{y}$

then $\exists\ \eta, \gamma > 0$ with $\eta < \frac{m^2}{L}$ such that $\forall\ \mathbf{w}$:

Damped phase:				
$\Phi(\mathbf{w}^+) - \Phi(\mathbf{w})$	$\leq$	$-\gamma$	if	$\|\nabla \Phi(\mathbf{w})\| \geq \eta$
Quadratic phase:				
$\|\nabla \Phi(\mathbf{w}^+)\|$	$\leq$	$\frac{L}{2m^2}\|\nabla \Phi(\mathbf{w})\|^2$	if	$\|\nabla \Phi(\mathbf{w})\| < \eta$

where $\mathbf{w}^+ = \mathbf{w} + t\Delta\mathbf{w}$ is the (possibly) reduced Newton step

Practical consequence: two-phase convergence

- If $\mathbf{w}$ is *far* from $\mathbf{w}^*$ $\Rightarrow$ Damped convergence (reduced steps)
- If $\mathbf{w}$ is *close* to $\mathbf{w}^*$ $\Rightarrow$ Quadratic convergence (full steps)
- Once Newton has entered the quadratic phase, it stays quadratic !!

Convergence of the Newton method with Line-search (unconstrained)

Theorem

Assume:

- Hessian is bounded $m \cdot I \leq \nabla^2 \Phi(\mathbf{w}) \leq M \cdot I$, with $m, M > 0$
- Hessian is Lipschitz, i.e. $\|\nabla^2 \Phi(\mathbf{w}) - \nabla^2 \Phi(\mathbf{y})\| \leq L\|\mathbf{w} - \mathbf{y}\|, \quad \forall \mathbf{w}, \mathbf{y}$

then $\exists\ \eta, \gamma > 0$ with $\eta < \frac{m^2}{L}$ such that $\forall\ \mathbf{w}$:

Damped phase:				
$\Phi(\mathbf{w}^+) - \Phi(\mathbf{w})$	$\leq$	$-\gamma$	if	$\|\nabla \Phi(\mathbf{w})\| \geq \eta$
Quadratic phase:				
$\|\nabla \Phi(\mathbf{w}^+)\|$	$\leq$	$\frac{L}{2m^2}\|\nabla \Phi(\mathbf{w})\|^2$	if	$\|\nabla \Phi(\mathbf{w})\| < \eta$

where $\mathbf{w}^+ = \mathbf{w} + t\Delta\mathbf{w}$ is the (possibly) reduced Newton step

Practical consequence: region of quadratic convergence

- If L is large, then
  1. the quadratic region ($\|\nabla \Phi(\mathbf{w})\| < \eta$) is small
  2. the quadratic contraction rate is small

Convergence of the Newton method with Line-search (unconstrained)

Theorem

Assume:

- Hessian is bounded $m \cdot I \leq \nabla^2 \Phi(\mathbf{w}) \leq M \cdot I$, with $m, M > 0$
- Hessian is Lipschitz, i.e. $\|\nabla^2 \Phi(\mathbf{w}) - \nabla^2 \Phi(\mathbf{y})\| \leq L\|\mathbf{w} - \mathbf{y}\|, \quad \forall \mathbf{w}, \mathbf{y}$

then $\exists\ \eta, \gamma > 0$ with $\eta < \frac{m^2}{L}$ such that $\forall\ \mathbf{w}$:

Damped phase:				
$\Phi(\mathbf{w}^+) - \Phi(\mathbf{w})$	$\leq$	$-\gamma$	if	$\|\nabla \Phi(\mathbf{w})\| \geq \eta$
Quadratic phase:				
$\|\nabla \Phi(\mathbf{w}^+)\|$	$\leq$	$\frac{L}{2m^2}\|\nabla \Phi(\mathbf{w})\|^2$	if	$\|\nabla \Phi(\mathbf{w})\| < \eta$

where $\mathbf{w}^+ = \mathbf{w} + t\Delta\mathbf{w}$ is the (possibly) reduced Newton step

What if we have constraints ? Similar results, with $\nabla^2_{\mathbf{w}} \mathcal{L}$, $\nabla_{\mathbf{w}} \mathbf{g}$. However, Newton can fail on an infeasible point (no direction improving the constraints, c.f. "But still, Newton can fail...")

Outline

1. The Newton method
2. Newton on the KKT conditions
3. The reduced Newton step (unconstrained problems)
4. The merit function - Line-search for constrained problems
5. Newton-type methods
6. Sequential Quadratic Programming

Assessing convergence with constraints

Solve the NLP...

$$\begin{aligned} \min_{\mathbf{w}} \quad & \Phi(\mathbf{w}) \\ \text{s.t.} \quad & \mathbf{g}(\mathbf{w}) = 0 \end{aligned}$$

Assessing convergence with constraints

Solve the NLP...

$$\begin{aligned} \min_{\mathbf{w}} \quad & \Phi(\mathbf{w}) \\ \text{s.t.} \quad & \mathbf{g}(\mathbf{w}) = 0 \end{aligned}$$

... by satisfying the KKT conditions...

$$\mathbf{r}(\mathbf{w}, \boldsymbol{\lambda}) = \begin{bmatrix} \nabla_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}) \\ \mathbf{g}(\mathbf{w}) \end{bmatrix} = 0$$

Assessing convergence with constraints

Solve the NLP...

$$\begin{aligned} \min_{\mathbf{w}} \quad & \Phi(\mathbf{w}) \\ \text{s.t.} \quad & \mathbf{g}(\mathbf{w}) = 0 \end{aligned}$$

... by satisfying the KKT conditions...

$$\mathbf{r}(\mathbf{w}, \boldsymbol{\lambda}) = \begin{bmatrix} \nabla_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}) \\ \mathbf{g}(\mathbf{w}) \end{bmatrix} = 0$$

... via taking Newton steps along the Newton directions.

Assessing convergence with constraints

Solve the NLP...

$$\begin{aligned} \min_{\mathbf{w}} \quad & \Phi(\mathbf{w}) \\ \text{s.t.} \quad & \mathbf{g}(\mathbf{w}) = 0 \end{aligned}$$

... by satisfying the KKT conditions...

$$\mathbf{r}(\mathbf{w}, \boldsymbol{\lambda}) = \begin{bmatrix} \nabla_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}) \\ \mathbf{g}(\mathbf{w}) \end{bmatrix} = 0$$

... via taking Newton steps along the Newton directions.

The Newton directions are given by

$$\begin{bmatrix} H(\mathbf{w}, \boldsymbol{\lambda}) & \nabla \mathbf{g}(\mathbf{w}) \\ \nabla \mathbf{g}(\mathbf{w})^{\mathsf{T}} & 0 \end{bmatrix} \begin{bmatrix} \Delta \mathbf{w} \\ \boldsymbol{\lambda}^{+} \end{bmatrix} = - \begin{bmatrix} \nabla \Phi(\mathbf{w}) \\ \mathbf{g}(\mathbf{w}) \end{bmatrix}$$

and $\Delta \boldsymbol{\lambda} = \boldsymbol{\lambda}^{+} - \boldsymbol{\lambda}$

Assessing convergence with constraints

Solve the NLP...

$$\begin{aligned} \min_{\mathbf{w}} \quad & \Phi(\mathbf{w}) \\ \text{s.t.} \quad & \mathbf{g}(\mathbf{w}) = 0 \end{aligned}$$

... by satisfying the KKT conditions...

$$\mathbf{r}(\mathbf{w}, \boldsymbol{\lambda}) = \begin{bmatrix} \nabla_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}) \\ \mathbf{g}(\mathbf{w}) \end{bmatrix} = 0$$

... via taking Newton steps along the Newton directions.

The Newton directions are given by

$$\begin{bmatrix} H(\mathbf{w}, \boldsymbol{\lambda}) & \nabla \mathbf{g}(\mathbf{w}) \\ \nabla \mathbf{g}(\mathbf{w})^T & 0 \end{bmatrix} \begin{bmatrix} \Delta \mathbf{w} \\ \boldsymbol{\lambda}^+ \end{bmatrix} = - \begin{bmatrix} \nabla \Phi(\mathbf{w}) \\ \mathbf{g}(\mathbf{w}) \end{bmatrix}$$

and $\Delta \boldsymbol{\lambda} = \boldsymbol{\lambda}^+ - \boldsymbol{\lambda}$

- We know that there exists a Newton step that improves $\|\mathbf{r}\|$, i.e.

$$\|\mathbf{r}(\mathbf{w} + t\Delta \mathbf{w}, \boldsymbol{\lambda} + t\Delta \boldsymbol{\lambda})\| < \|\mathbf{r}(\mathbf{w}, \boldsymbol{\lambda})\|$$

  for some $t \in]0, 1]$ (as long as $\nabla \mathbf{r}$ is non-singular). How to select t ?

Assessing convergence with constraints

Solve the NLP... ... by satisfying the KKT conditions...

$$\begin{aligned} \min_{\mathbf{w}} \quad & \Phi(\mathbf{w}) \\ \text{s.t.} \quad & \mathbf{g}(\mathbf{w}) = 0 \end{aligned}$$

$$\mathbf{r}(\mathbf{w}, \boldsymbol{\lambda}) = \begin{bmatrix} \nabla_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}) \\ \mathbf{g}(\mathbf{w}) \end{bmatrix} = 0$$

... via taking Newton steps along the Newton directions.

The Newton directions are given by

$$\begin{bmatrix} H(\mathbf{w}, \boldsymbol{\lambda}) & \nabla \mathbf{g}(\mathbf{w}) \\ \nabla \mathbf{g}(\mathbf{w})^T & 0 \end{bmatrix} \begin{bmatrix} \Delta \mathbf{w} \\ \boldsymbol{\lambda}^+ \end{bmatrix} = - \begin{bmatrix} \nabla \Phi(\mathbf{w}) \\ \mathbf{g}(\mathbf{w}) \end{bmatrix}$$

and $\Delta \boldsymbol{\lambda} = \boldsymbol{\lambda}^+ - \boldsymbol{\lambda}$

- We know that there exists a Newton step that improves $\|\mathbf{r}\|$, i.e.

$$\|\mathbf{r}(\mathbf{w} + t\Delta \mathbf{w}, \boldsymbol{\lambda} + t\Delta \boldsymbol{\lambda})\| < \|\mathbf{r}(\mathbf{w}, \boldsymbol{\lambda})\|$$

  for some $t \in]0, 1]$ (as long as $\nabla \mathbf{r}$ is non-singular). How to select t ?

- Could ensure the progress of $\|\mathbf{r}(\mathbf{w} + t\Delta \mathbf{w}, \boldsymbol{\lambda} + t\Delta \boldsymbol{\lambda})\|$

Assessing convergence with constraints

Solve the NLP...

$$\begin{aligned} \min_{\mathbf{w}} \quad & \Phi(\mathbf{w}) \\ \text{s.t.} \quad & \mathbf{g}(\mathbf{w}) = 0 \end{aligned}$$

... by satisfying the KKT conditions...

$$\mathbf{r}(\mathbf{w}, \boldsymbol{\lambda}) = \begin{bmatrix} \nabla_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}) \\ \mathbf{g}(\mathbf{w}) \end{bmatrix} = 0$$

... via taking Newton steps along the Newton directions.

The Newton directions are given by

$$\begin{bmatrix} H(\mathbf{w}, \boldsymbol{\lambda}) & \nabla \mathbf{g}(\mathbf{w}) \\ \nabla \mathbf{g}(\mathbf{w})^T & 0 \end{bmatrix} \begin{bmatrix} \Delta \mathbf{w} \\ \boldsymbol{\lambda}^+ \end{bmatrix} = - \begin{bmatrix} \nabla \Phi(\mathbf{w}) \\ \mathbf{g}(\mathbf{w}) \end{bmatrix}$$

and $\Delta \boldsymbol{\lambda} = \boldsymbol{\lambda}^+ - \boldsymbol{\lambda}$

- We know that there exists a Newton step that improves $\|\mathbf{r}\|$, i.e.

$$\|\mathbf{r}(\mathbf{w} + t\Delta \mathbf{w}, \boldsymbol{\lambda} + t\Delta \boldsymbol{\lambda})\| < \|\mathbf{r}(\mathbf{w}, \boldsymbol{\lambda})\|$$

  for some $t \in]0, 1]$ (as long as $\nabla \mathbf{r}$ is non-singular). How to select t ?
- Could ensure the progress of $\|\mathbf{r}(\mathbf{w} + t\Delta \mathbf{w}, \boldsymbol{\lambda} + t\Delta \boldsymbol{\lambda})\|$
- Some caveats... e.g. need to evaluate $\nabla \mathbf{g}(\mathbf{w} + t\Delta \mathbf{w})$ each time we backtrack on t !! This can be computationally expensive...

Line-search with Equality Constraints

T_1 merit function...

... accounts for progress on Φ and $\mathbf{g}$

$$\boxed{T_1(\mathbf{w}) = \Phi(\mathbf{w}) + \nu \|\mathbf{g}(\mathbf{w})\|_1}$$

We would like $\min_{\mathbf{w}} T_1(\mathbf{w})$ to coincide with the solution of the NLP !!

Line-search with Equality Constraints

T_1 merit function...

... accounts for progress on Φ and $\mathbf{g}$

$$\boxed{T_1(\mathbf{w}) = \Phi(\mathbf{w}) + \nu \|\mathbf{g}(\mathbf{w})\|_1}$$

We would like $\min_{\mathbf{w}} T_1(\mathbf{w})$ to coincide with the solution of the NLP !!

Properties of the T_1 merit function:

- If $\mathbf{w}^*, \boldsymbol{\lambda}^*$ is a regular KKT point with SOSC, then $\mathbf{w}^*$ is a (local) minimum of T_1 for $\nu > \|\boldsymbol{\lambda}^*\|_\infty$

Line-search with Equality Constraints

T_1 merit function...

... accounts for progress on Φ and $\mathbf{g}$

$$\boxed{T_1(\mathbf{w}) = \Phi(\mathbf{w}) + \nu \|\mathbf{g}(\mathbf{w})\|_1}$$

We would like $\min_{\mathbf{w}} T_1(\mathbf{w})$ to coincide with the solution of the NLP !!

Properties of the T_1 merit function:

- If $\mathbf{w}^*, \boldsymbol{\lambda}^*$ is a regular KKT point with SOSC, then $\mathbf{w}^*$ is a (local) minimum of T_1 for $\nu > \|\boldsymbol{\lambda}^*\|_\infty$
- If $\mathbf{w}$ is a minimum of T_1 for $\nu > \|\boldsymbol{\lambda}^*\|_\infty$ with $\mathbf{g}(\mathbf{w}) = 0$, then $\mathbf{w}$ is a KKT point of the original problem. If $\mathbf{g}(\mathbf{w}) \neq 0$, then we have collapsed to an infeasible point.

Line-search with Equality Constraints

T_1 merit function...

... accounts for progress on Φ and $\mathbf{g}$

$$\boxed{T_1(\mathbf{w}) = \Phi(\mathbf{w}) + \nu \|\mathbf{g}(\mathbf{w})\|_1}$$

We would like $\min_{\mathbf{w}} T_1(\mathbf{w})$ to coincide with the solution of the NLP !!

Properties of the T_1 merit function:

- If $\mathbf{w}^*, \boldsymbol{\lambda}^*$ is a regular KKT point with SOSC, then $\mathbf{w}^*$ is a (local) minimum of T_1 for $\nu > \|\boldsymbol{\lambda}^*\|_\infty$
- If $\mathbf{w}$ is a minimum of T_1 for $\nu > \|\boldsymbol{\lambda}^*\|_\infty$ with $\mathbf{g}(\mathbf{w}) = 0$, then $\mathbf{w}$ is a KKT point of the original problem. If $\mathbf{g}(\mathbf{w}) \neq 0$, then we have collapsed to an infeasible point.

Merit function T_1 cannot be improved from $\mathbf{w}^*$. Minimising T_1 yields a KKT point.

Line-search with Equality Constraints

T_1 merit function...

... accounts for progress on Φ and $\mathbf{g}$

$$\boxed{T_1(\mathbf{w}) = \Phi(\mathbf{w}) + \nu \|\mathbf{g}(\mathbf{w})\|_1}$$

We would like $\min_{\mathbf{w}} T_1(\mathbf{w})$ to coincide with the solution of the NLP !!

Properties of the T_1 merit function:

- If $\mathbf{w}^*, \boldsymbol{\lambda}^*$ is a regular KKT point with SOSC, then $\mathbf{w}^*$ is a (local) minimum of T_1 for $\nu > \|\boldsymbol{\lambda}^*\|_\infty$
- If $\mathbf{w}$ is a minimum of T_1 for $\nu > \|\boldsymbol{\lambda}^*\|_\infty$ with $\mathbf{g}(\mathbf{w}) = 0$, then $\mathbf{w}$ is a KKT point of the original problem. If $\mathbf{g}(\mathbf{w}) \neq 0$, then we have collapsed to an infeasible point.

Merit function T_1 cannot be improved from $\mathbf{w}^*$. Minimising T_1 yields a KKT point.

What about convergence of Newton with line-search on T_1 ?

Similar results as in the unconstrained case if LICQ is respected !!

Line-search with Equality Constraints

T_1 merit function...

... accounts for progress on Φ and $\mathbf{g}$

$$\boxed{T_1(\mathbf{w}) = \Phi(\mathbf{w}) + \nu \|\mathbf{g}(\mathbf{w})\|_1}$$

We would like $\min_{\mathbf{w}} T_1(\mathbf{w})$ to coincide with the solution of the NLP !!

Properties of the T_1 merit function:

- If $\mathbf{w}^*, \boldsymbol{\lambda}^*$ is a regular KKT point with SOSC, then $\mathbf{w}^*$ is a (local) minimum of T_1 for $\nu > \|\boldsymbol{\lambda}^*\|_\infty$
- If $\mathbf{w}$ is a minimum of T_1 for $\nu > \|\boldsymbol{\lambda}^*\|_\infty$ with $\mathbf{g}(\mathbf{w}) = 0$, then $\mathbf{w}$ is a KKT point of the original problem. If $\mathbf{g}(\mathbf{w}) \neq 0$, then we have collapsed to an infeasible point.

Merit function T_1 cannot be improved from $\mathbf{w}^*$. Minimising T_1 yields a KKT point.

What about convergence of Newton with line-search on T_1 ?

Similar results as in the unconstrained case if LICQ is respected !!

Difficulty: $\boldsymbol{\lambda}^*$ is not know a priori. Parameter ν must be chosen adequately, possibly adjusted over the iterations !

Newton direction is a descent direction for T_1

Theorem: for $\mathbf{g}_i(\mathbf{w}) \neq 0 \ \forall i$, then the Newton direction $(\mathbf{w}, \boldsymbol{\lambda})$

$$\begin{bmatrix} H(\mathbf{w}, \boldsymbol{\lambda}) & \nabla \mathbf{g}(\mathbf{w}) \\ \nabla \mathbf{g}(\mathbf{w})^\top & 0 \end{bmatrix} \begin{bmatrix} \Delta \mathbf{w} \\ \boldsymbol{\lambda} \end{bmatrix} = - \begin{bmatrix} \nabla \Phi(\mathbf{w}) \\ \mathbf{g}(\mathbf{w}) \end{bmatrix}$$

is a descent direction for the T_1 merit function $T_1 = \Phi(\mathbf{w}) + \nu \|\mathbf{g}(\mathbf{w})\|_1$.

Newton direction is a descent direction for T_1

Theorem: for $\mathbf{g}_i(\mathbf{w}) \neq 0 \ \forall i$, then the Newton direction $(\mathbf{w}, \boldsymbol{\lambda})$

$$\begin{bmatrix} H(\mathbf{w}, \boldsymbol{\lambda}) & \nabla \mathbf{g}(\mathbf{w}) \\ \nabla \mathbf{g}(\mathbf{w})^\top & 0 \end{bmatrix} \begin{bmatrix} \Delta \mathbf{w} \\ \boldsymbol{\lambda} \end{bmatrix} = - \begin{bmatrix} \nabla \Phi(\mathbf{w}) \\ \mathbf{g}(\mathbf{w}) \end{bmatrix}$$

is a descent direction for the T_1 merit function $T_1 = \Phi(\mathbf{w}) + \nu \|\mathbf{g}(\mathbf{w})\|_1$.

Proof: $\Delta \mathbf{w}$ is a descent direction if

$$\frac{\mathrm{d}}{\mathrm{d}t} T_1(\mathbf{w} + t\Delta \mathbf{w})\Big|_{t=0} < 0$$

Newton direction is a descent direction for T_1

Theorem: for $\mathbf{g}_i(\mathbf{w}) \neq 0 \ \forall i$, then the Newton direction $(\mathbf{w}, \boldsymbol{\lambda})$

$$\begin{bmatrix} H(\mathbf{w}, \boldsymbol{\lambda}) & \nabla \mathbf{g}(\mathbf{w}) \\ \nabla \mathbf{g}(\mathbf{w})^\top & 0 \end{bmatrix} \begin{bmatrix} \Delta \mathbf{w} \\ \boldsymbol{\lambda} \end{bmatrix} = - \begin{bmatrix} \nabla \Phi(\mathbf{w}) \\ \mathbf{g}(\mathbf{w}) \end{bmatrix}$$

is a descent direction for the T_1 merit function $T_1 = \Phi(\mathbf{w}) + \nu \|\mathbf{g}(\mathbf{w})\|_1$.

Proof: $\Delta \mathbf{w}$ is a descent direction if

$$\frac{\mathrm{d}}{\mathrm{d}t} T_1(\mathbf{w} + t\Delta \mathbf{w})\Big|_{t=0} < 0$$

For $\mathbf{g}_i(\mathbf{w}) \neq 0 \ \forall i$, using $\boldsymbol{\sigma} = \operatorname{sign}(\mathbf{g}(\mathbf{w}))$ (i.e. $\boldsymbol{\sigma}^\top \mathbf{g}(\mathbf{w}) = \|\mathbf{g}(\mathbf{w})\|_1$)

Newton direction is a descent direction for T_1

Theorem: for $\mathbf{g}_i(\mathbf{w}) \neq 0 \ \forall i$, then the Newton direction $(\mathbf{w}, \boldsymbol{\lambda})$

$$\begin{bmatrix} H(\mathbf{w}, \boldsymbol{\lambda}) & \nabla \mathbf{g}(\mathbf{w}) \\ \nabla \mathbf{g}(\mathbf{w})^\top & 0 \end{bmatrix} \begin{bmatrix} \Delta \mathbf{w} \\ \boldsymbol{\lambda} \end{bmatrix} = - \begin{bmatrix} \nabla \Phi(\mathbf{w}) \\ \mathbf{g}(\mathbf{w}) \end{bmatrix}$$

is a descent direction for the T_1 merit function $T_1 = \Phi(\mathbf{w}) + \nu \|\mathbf{g}(\mathbf{w})\|_1$.

Proof: $\Delta \mathbf{w}$ is a descent direction if

$$\frac{\mathrm{d}}{\mathrm{d}t} T_1(\mathbf{w} + t\Delta \mathbf{w})\Big|_{t=0} < 0$$

For $\mathbf{g}_i(\mathbf{w}) \neq 0 \ \forall i$, using $\boldsymbol{\sigma} = \operatorname{sign}(\mathbf{g}(\mathbf{w}))$ (i.e. $\boldsymbol{\sigma}^\top \mathbf{g}(\mathbf{w}) = \|\mathbf{g}(\mathbf{w})\|_1$)

$$\frac{\mathrm{d}}{\mathrm{d}t} \|\mathbf{g}(\mathbf{w} + t\Delta \mathbf{w})\|_1 \big|_{t=0} = \boldsymbol{\sigma}^\top \nabla \mathbf{g}(\mathbf{w})^\top \Delta \mathbf{w}$$

Newton direction is a descent direction for T_1

Theorem: for $\mathbf{g}_i(\mathbf{w}) \neq 0 \ \forall i$, then the Newton direction $(\mathbf{w}, \boldsymbol{\lambda})$

$$\begin{bmatrix} H(\mathbf{w}, \boldsymbol{\lambda}) & \nabla \mathbf{g}(\mathbf{w}) \\ \nabla \mathbf{g}(\mathbf{w})^\top & 0 \end{bmatrix} \begin{bmatrix} \Delta \mathbf{w} \\ \boldsymbol{\lambda} \end{bmatrix} = - \begin{bmatrix} \nabla \Phi(\mathbf{w}) \\ \mathbf{g}(\mathbf{w}) \end{bmatrix}$$

is a descent direction for the T_1 merit function $T_1 = \Phi(\mathbf{w}) + \nu \|\mathbf{g}(\mathbf{w})\|_1$.

Proof: $\Delta \mathbf{w}$ is a descent direction if

$$\frac{\mathrm{d}}{\mathrm{d}t} T_1(\mathbf{w} + t\Delta \mathbf{w})\Big|_{t=0} < 0$$

For $\mathbf{g}_i(\mathbf{w}) \neq 0 \ \forall i$, using $\boldsymbol{\sigma} = \operatorname{sign}(\mathbf{g}(\mathbf{w}))$ (i.e. $\boldsymbol{\sigma}^\top \mathbf{g}(\mathbf{w}) = \|\mathbf{g}(\mathbf{w})\|_1$)

$$\frac{\mathrm{d}}{\mathrm{d}t} \|\mathbf{g}(\mathbf{w} + t\Delta \mathbf{w})\|_1 \big|_{t=0} = \boldsymbol{\sigma}^\top \nabla \mathbf{g}(\mathbf{w})^\top \Delta \mathbf{w}$$

and since $\nabla \mathbf{g}(\mathbf{w})^\top \Delta \mathbf{w} = -\mathbf{g}(\mathbf{w})$, we have:

$$\frac{\mathrm{d}}{\mathrm{d}t} \|\mathbf{g}(\mathbf{w} + t\Delta \mathbf{w})\|_{t=0} = -\boldsymbol{\sigma}^\top \mathbf{g}(\mathbf{w}) = -\|\mathbf{g}(\mathbf{w})\|_1$$

Newton direction is a descent direction for T_1

Theorem: for $\mathbf{g}_i(\mathbf{w}) \neq 0 \ \forall i$, then the Newton direction $(\mathbf{w}, \boldsymbol{\lambda})$

$$\begin{bmatrix} H(\mathbf{w}, \boldsymbol{\lambda}) & \nabla \mathbf{g}(\mathbf{w}) \\ \nabla \mathbf{g}(\mathbf{w})^\top & 0 \end{bmatrix} \begin{bmatrix} \Delta \mathbf{w} \\ \boldsymbol{\lambda} \end{bmatrix} = - \begin{bmatrix} \nabla \Phi(\mathbf{w}) \\ \mathbf{g}(\mathbf{w}) \end{bmatrix}$$

is a descent direction for the T_1 merit function $T_1 = \Phi(\mathbf{w}) + \nu \|\mathbf{g}(\mathbf{w})\|_1$.

Proof: $\Delta \mathbf{w}$ is a descent direction if

$$\frac{\mathrm{d}}{\mathrm{d}t} T_1(\mathbf{w} + t\Delta \mathbf{w})\Big|_{t=0} < 0$$

We have:

$$\frac{\mathrm{d}}{\mathrm{d}t} \|\mathbf{g}(\mathbf{w} + t\Delta \mathbf{w})\|_{t=0} = -\boldsymbol{\sigma}^\top \mathbf{g}(\mathbf{w}) = -\|\mathbf{g}(\mathbf{w})\|_1$$

Newton direction is a descent direction for T_1

Theorem: for $\mathbf{g}_i(\mathbf{w}) \neq 0 \ \forall i$, then the Newton direction $(\mathbf{w}, \boldsymbol{\lambda})$

$$\begin{bmatrix} H(\mathbf{w}, \boldsymbol{\lambda}) & \nabla \mathbf{g}(\mathbf{w}) \\ \nabla \mathbf{g}(\mathbf{w})^\top & 0 \end{bmatrix} \begin{bmatrix} \Delta \mathbf{w} \\ \boldsymbol{\lambda} \end{bmatrix} = - \begin{bmatrix} \nabla \Phi(\mathbf{w}) \\ \mathbf{g}(\mathbf{w}) \end{bmatrix}$$

is a descent direction for the T_1 merit function $T_1 = \Phi(\mathbf{w}) + \nu \|\mathbf{g}(\mathbf{w})\|_1$.

Proof: $\Delta \mathbf{w}$ is a descent direction if

$$\frac{\mathrm{d}}{\mathrm{d}t} T_1(\mathbf{w} + t\Delta \mathbf{w})\Big|_{t=0} < 0$$

We have:

$$\frac{\mathrm{d}}{\mathrm{d}t} \|\mathbf{g}(\mathbf{w} + t\Delta \mathbf{w})\|_{t=0} = -\boldsymbol{\sigma}^\top \mathbf{g}(\mathbf{w}) = -\|\mathbf{g}(\mathbf{w})\|_1$$

Moreover, $H\Delta \mathbf{w} + \nabla \mathbf{g} \boldsymbol{\lambda} = -\nabla \Phi$ yields $\Delta \mathbf{w}^\top H \Delta \mathbf{w} + \Delta \mathbf{w}^\top \nabla \mathbf{g} \boldsymbol{\lambda} = -\Delta \mathbf{w}^\top \nabla \Phi$

Newton direction is a descent direction for T_1

Theorem: for $\mathbf{g}_i(\mathbf{w}) \neq 0 \ \forall i$, then the Newton direction $(\mathbf{w}, \boldsymbol{\lambda})$

$$\begin{bmatrix} H(\mathbf{w}, \boldsymbol{\lambda}) & \nabla \mathbf{g}(\mathbf{w}) \\ \nabla \mathbf{g}(\mathbf{w})^\top & 0 \end{bmatrix} \begin{bmatrix} \Delta \mathbf{w} \\ \boldsymbol{\lambda} \end{bmatrix} = - \begin{bmatrix} \nabla \Phi(\mathbf{w}) \\ \mathbf{g}(\mathbf{w}) \end{bmatrix}$$

is a descent direction for the T_1 merit function $T_1 = \Phi(\mathbf{w}) + \nu \|\mathbf{g}(\mathbf{w})\|_1$.

Proof: $\Delta \mathbf{w}$ is a descent direction if

$$\frac{\mathrm{d}}{\mathrm{d}t} T_1(\mathbf{w} + t\Delta \mathbf{w})\Big|_{t=0} < 0$$

We have:

$$\frac{\mathrm{d}}{\mathrm{d}t} \|\mathbf{g}(\mathbf{w} + t\Delta \mathbf{w})\|_{t=0} = -\boldsymbol{\sigma}^\top \mathbf{g}(\mathbf{w}) = -\|\mathbf{g}(\mathbf{w})\|_1$$

Moreover, $H\Delta \mathbf{w} + \nabla \mathbf{g} \boldsymbol{\lambda} = -\nabla \Phi$ yields $\Delta \mathbf{w}^\top H \Delta \mathbf{w} + \Delta \mathbf{w}^\top \nabla \mathbf{g} \boldsymbol{\lambda} = -\Delta \mathbf{w}^\top \nabla \Phi$, and:

$$\frac{\mathrm{d}}{\mathrm{d}t} \Phi(\mathbf{w} + t\Delta \mathbf{w})\Big|_{t=0} = \nabla \Phi^\top \Delta \mathbf{w}$$

Newton direction is a descent direction for T_1

Theorem: for $\mathbf{g}_i(\mathbf{w}) \neq 0 \ \forall i$, then the Newton direction $(\mathbf{w}, \boldsymbol{\lambda})$

$$\begin{bmatrix} H(\mathbf{w}, \boldsymbol{\lambda}) & \nabla \mathbf{g}(\mathbf{w}) \\ \nabla \mathbf{g}(\mathbf{w})^\top & 0 \end{bmatrix} \begin{bmatrix} \Delta \mathbf{w} \\ \boldsymbol{\lambda} \end{bmatrix} = - \begin{bmatrix} \nabla \Phi(\mathbf{w}) \\ \mathbf{g}(\mathbf{w}) \end{bmatrix}$$

is a descent direction for the T_1 merit function $T_1 = \Phi(\mathbf{w}) + \nu \|\mathbf{g}(\mathbf{w})\|_1$.

Proof: $\Delta \mathbf{w}$ is a descent direction if

$$\frac{\mathrm{d}}{\mathrm{d}t} T_1(\mathbf{w} + t\Delta \mathbf{w})\Big|_{t=0} < 0$$

We have:

$$\frac{\mathrm{d}}{\mathrm{d}t} \|\mathbf{g}(\mathbf{w} + t\Delta \mathbf{w})\|_{t=0} = -\boldsymbol{\sigma}^\top \mathbf{g}(\mathbf{w}) = -\|\mathbf{g}(\mathbf{w})\|_1$$

Moreover, $H\Delta \mathbf{w} + \nabla \mathbf{g} \boldsymbol{\lambda} = -\nabla \Phi$ yields $\Delta \mathbf{w}^\top H \Delta \mathbf{w} + \Delta \mathbf{w}^\top \nabla \mathbf{g} \boldsymbol{\lambda} = -\Delta \mathbf{w}^\top \nabla \Phi$, and:

$$\frac{\mathrm{d}}{\mathrm{d}t} \Phi(\mathbf{w} + t\Delta \mathbf{w})\Big|_{t=0} = \nabla \Phi^\top \Delta \mathbf{w} = -\Delta \mathbf{w}^\top H \Delta \mathbf{w} - \boldsymbol{\lambda} \nabla \mathbf{g}^\top \Delta \mathbf{w}$$

Newton direction is a descent direction for T_1

Theorem: for $\mathbf{g}_i(\mathbf{w}) \neq 0 \ \forall i$, then the Newton direction $(\mathbf{w}, \boldsymbol{\lambda})$

$$\begin{bmatrix} H(\mathbf{w}, \boldsymbol{\lambda}) & \nabla \mathbf{g}(\mathbf{w}) \\ \nabla \mathbf{g}(\mathbf{w})^\top & 0 \end{bmatrix} \begin{bmatrix} \Delta \mathbf{w} \\ \boldsymbol{\lambda} \end{bmatrix} = - \begin{bmatrix} \nabla \Phi(\mathbf{w}) \\ \mathbf{g}(\mathbf{w}) \end{bmatrix}$$

is a descent direction for the T_1 merit function $T_1 = \Phi(\mathbf{w}) + \nu \|\mathbf{g}(\mathbf{w})\|_1$.

Proof: $\Delta \mathbf{w}$ is a descent direction if

$$\frac{\mathrm{d}}{\mathrm{d}t} T_1(\mathbf{w} + t\Delta \mathbf{w})\Big|_{t=0} < 0$$

We have:

$$\frac{\mathrm{d}}{\mathrm{d}t} \|\mathbf{g}(\mathbf{w} + t\Delta \mathbf{w})\|_{t=0} = -\boldsymbol{\sigma}^\top \mathbf{g}(\mathbf{w}) = -\|\mathbf{g}(\mathbf{w})\|_1$$

Moreover, $H\Delta \mathbf{w} + \nabla \mathbf{g} \boldsymbol{\lambda} = -\nabla \Phi$ yields $\Delta \mathbf{w}^\top H \Delta \mathbf{w} + \Delta \mathbf{w}^\top \nabla \mathbf{g} \boldsymbol{\lambda} = -\Delta \mathbf{w}^\top \nabla \Phi$, and:

$$\frac{\mathrm{d}}{\mathrm{d}t} \Phi(\mathbf{w} + t\Delta \mathbf{w})\Big|_{t=0} = \nabla \Phi^\top \Delta \mathbf{w} = -\Delta \mathbf{w}^\top H \Delta \mathbf{w} - \boldsymbol{\lambda} \nabla \mathbf{g}^\top \Delta \mathbf{w}$$

Such that:

$$\frac{\mathrm{d}}{\mathrm{d}t} T_1(\mathbf{w} + t\Delta \mathbf{w})\Big|_{t=0} = -\Delta \mathbf{w}^\top H \Delta \mathbf{w} - \boldsymbol{\lambda} \nabla \mathbf{g}^\top \Delta \mathbf{w} - \nu \|\mathbf{g}(\mathbf{w})\|_1$$

Newton direction is a descent direction for T_1

Theorem: for $\mathbf{g}_i(\mathbf{w}) \neq 0 \ \forall i$, then the Newton direction $(\mathbf{w}, \boldsymbol{\lambda})$

$$\begin{bmatrix} H(\mathbf{w}, \boldsymbol{\lambda}) & \nabla \mathbf{g}(\mathbf{w}) \\ \nabla \mathbf{g}(\mathbf{w})^\top & 0 \end{bmatrix} \begin{bmatrix} \Delta \mathbf{w} \\ \boldsymbol{\lambda} \end{bmatrix} = - \begin{bmatrix} \nabla \Phi(\mathbf{w}) \\ \mathbf{g}(\mathbf{w}) \end{bmatrix}$$

is a descent direction for the T_1 merit function $T_1 = \Phi(\mathbf{w}) + \nu \|\mathbf{g}(\mathbf{w})\|_1$.

Proof: $\Delta \mathbf{w}$ is a descent direction if

$$\frac{\mathrm{d}}{\mathrm{d}t} T_1(\mathbf{w} + t\Delta \mathbf{w})\Big|_{t=0} < 0$$

We have:

$$\frac{\mathrm{d}}{\mathrm{d}t} \|\mathbf{g}(\mathbf{w} + t\Delta \mathbf{w})\|_{t=0} = -\boldsymbol{\sigma}^\top \mathbf{g}(\mathbf{w}) = -\|\mathbf{g}(\mathbf{w})\|_1$$

Moreover, $H\Delta \mathbf{w} + \nabla \mathbf{g} \boldsymbol{\lambda} = -\nabla \Phi$ yields $\Delta \mathbf{w}^\top H \Delta \mathbf{w} + \Delta \mathbf{w}^\top \nabla \mathbf{g} \boldsymbol{\lambda} = -\Delta \mathbf{w}^\top \nabla \Phi$, and:

$$\frac{\mathrm{d}}{\mathrm{d}t} \Phi(\mathbf{w} + t\Delta \mathbf{w})\Big|_{t=0} = \nabla \Phi^\top \Delta \mathbf{w} = -\Delta \mathbf{w}^\top H \Delta \mathbf{w} - \boldsymbol{\lambda} \nabla \mathbf{g}^\top \Delta \mathbf{w}$$

Such that:

$$\frac{\mathrm{d}}{\mathrm{d}t} T_1(\mathbf{w} + t\Delta \mathbf{w})\Big|_{t=0} = -\Delta \mathbf{w}^\top H \Delta \mathbf{w} - \boldsymbol{\lambda} \nabla \mathbf{g}^\top \Delta \mathbf{w} - \nu \|\mathbf{g}(\mathbf{w})\|_1 =$$

$$-\Delta \mathbf{w}^\top H \Delta \mathbf{w} + \boldsymbol{\lambda} \mathbf{g} - \nu \|\mathbf{g}(\mathbf{w})\|_1$$

Newton direction is a descent direction for T_1

Theorem: for $\mathbf{g}_i(\mathbf{w}) \neq 0 \ \forall i$, then the Newton direction $(\mathbf{w}, \boldsymbol{\lambda})$

$$\begin{bmatrix} H(\mathbf{w}, \boldsymbol{\lambda}) & \nabla \mathbf{g}(\mathbf{w}) \\ \nabla \mathbf{g}(\mathbf{w})^\top & 0 \end{bmatrix} \begin{bmatrix} \Delta \mathbf{w} \\ \boldsymbol{\lambda} \end{bmatrix} = - \begin{bmatrix} \nabla \Phi(\mathbf{w}) \\ \mathbf{g}(\mathbf{w}) \end{bmatrix}$$

is a descent direction for the T_1 merit function $T_1 = \Phi(\mathbf{w}) + \nu \|\mathbf{g}(\mathbf{w})\|_1$.

Proof: $\Delta \mathbf{w}$ is a descent direction if

$$\frac{\mathrm{d}}{\mathrm{d}t} T_1(\mathbf{w} + t\Delta \mathbf{w})\Big|_{t=0} < 0$$

We have:

$$\frac{\mathrm{d}}{\mathrm{d}t} T_1(\mathbf{w} + t\Delta \mathbf{w})\Big|_{t=0} = -\Delta \mathbf{w}^\top H \Delta \mathbf{w} + \boldsymbol{\lambda} \mathbf{g} - \nu \|\mathbf{g}(\mathbf{w})\|_1$$

.

Newton direction is a descent direction for T_1

Theorem: for $\mathbf{g}_i(\mathbf{w}) \neq 0 \ \forall i$, then the Newton direction $(\mathbf{w}, \boldsymbol{\lambda})$

$$\begin{bmatrix} H(\mathbf{w}, \boldsymbol{\lambda}) & \nabla \mathbf{g}(\mathbf{w}) \\ \nabla \mathbf{g}(\mathbf{w})^\top & 0 \end{bmatrix} \begin{bmatrix} \Delta \mathbf{w} \\ \boldsymbol{\lambda} \end{bmatrix} = - \begin{bmatrix} \nabla \Phi(\mathbf{w}) \\ \mathbf{g}(\mathbf{w}) \end{bmatrix}$$

is a descent direction for the T_1 merit function $T_1 = \Phi(\mathbf{w}) + \nu \|\mathbf{g}(\mathbf{w})\|_1$.

Proof: $\Delta \mathbf{w}$ is a descent direction if

$$\frac{\mathrm{d}}{\mathrm{d}t} T_1(\mathbf{w} + t\Delta \mathbf{w})\Big|_{t=0} < 0$$

We have:

$$\frac{\mathrm{d}}{\mathrm{d}t} T_1(\mathbf{w} + t\Delta \mathbf{w})\Big|_{t=0} = -\Delta \mathbf{w}^\top H \Delta \mathbf{w} + \boldsymbol{\lambda} \mathbf{g} - \nu \|\mathbf{g}(\mathbf{w})\|_1$$

Then, since $|\boldsymbol{\lambda} \mathbf{g}| \leq \|\boldsymbol{\lambda}\|_\infty \|\mathbf{g}\|_1$, we have:

.

Newton direction is a descent direction for T_1

Theorem: for $\mathbf{g}_i(\mathbf{w}) \neq 0 \ \forall i$, then the Newton direction $(\mathbf{w}, \boldsymbol{\lambda})$

$$\begin{bmatrix} H(\mathbf{w}, \boldsymbol{\lambda}) & \nabla \mathbf{g}(\mathbf{w}) \\ \nabla \mathbf{g}(\mathbf{w})^\top & 0 \end{bmatrix} \begin{bmatrix} \Delta \mathbf{w} \\ \boldsymbol{\lambda} \end{bmatrix} = - \begin{bmatrix} \nabla \Phi(\mathbf{w}) \\ \mathbf{g}(\mathbf{w}) \end{bmatrix}$$

is a descent direction for the T_1 merit function $T_1 = \Phi(\mathbf{w}) + \nu \|\mathbf{g}(\mathbf{w})\|_1$.

Proof: $\Delta \mathbf{w}$ is a descent direction if

$$\frac{\mathrm{d}}{\mathrm{d}t} T_1(\mathbf{w} + t\Delta \mathbf{w})\Big|_{t=0} < 0$$

We have:

$$\frac{\mathrm{d}}{\mathrm{d}t} T_1(\mathbf{w} + t\Delta \mathbf{w})\Big|_{t=0} = -\Delta \mathbf{w}^\top H \Delta \mathbf{w} + \boldsymbol{\lambda} \mathbf{g} - \nu \|\mathbf{g}(\mathbf{w})\|_1$$

Then, since $|\boldsymbol{\lambda} \mathbf{g}| \leq \|\boldsymbol{\lambda}\|_\infty \|\mathbf{g}\|_1$, we have:

$$\frac{\mathrm{d}}{\mathrm{d}t} T_1(\mathbf{w} + t\Delta \mathbf{w})\Big|_{t=0} \leq -\Delta \mathbf{w}^\top H \Delta \mathbf{w} + \|\boldsymbol{\lambda}\|_\infty \|\mathbf{g}\|_1 - \nu \|\mathbf{g}(\mathbf{w})\|_1$$

.

Newton direction is a descent direction for T_1

Theorem: for $\mathbf{g}_i(\mathbf{w}) \neq 0 \ \forall i$, then the Newton direction $(\mathbf{w}, \boldsymbol{\lambda})$

$$\begin{bmatrix} H(\mathbf{w}, \boldsymbol{\lambda}) & \nabla \mathbf{g}(\mathbf{w}) \\ \nabla \mathbf{g}(\mathbf{w})^\top & 0 \end{bmatrix} \begin{bmatrix} \Delta \mathbf{w} \\ \boldsymbol{\lambda} \end{bmatrix} = - \begin{bmatrix} \nabla \Phi(\mathbf{w}) \\ \mathbf{g}(\mathbf{w}) \end{bmatrix}$$

is a descent direction for the T_1 merit function $T_1 = \Phi(\mathbf{w}) + \nu \|\mathbf{g}(\mathbf{w})\|_1$.

Proof: $\Delta \mathbf{w}$ is a descent direction if

$$\frac{\mathrm{d}}{\mathrm{d}t} T_1(\mathbf{w} + t\Delta \mathbf{w})\Big|_{t=0} < 0$$

We have:

$$\frac{\mathrm{d}}{\mathrm{d}t} T_1(\mathbf{w} + t\Delta \mathbf{w})\Big|_{t=0} = -\Delta \mathbf{w}^\top H \Delta \mathbf{w} + \boldsymbol{\lambda} \mathbf{g} - \nu \|\mathbf{g}(\mathbf{w})\|_1$$

Then, since $|\boldsymbol{\lambda} \mathbf{g}| \leq \|\boldsymbol{\lambda}\|_\infty \|\mathbf{g}\|_1$, we have:

$$\frac{\mathrm{d}}{\mathrm{d}t} T_1(\mathbf{w} + t\Delta \mathbf{w})\Big|_{t=0} \leq -\Delta \mathbf{w}^\top H \Delta \mathbf{w} + \|\boldsymbol{\lambda}\|_\infty \|\mathbf{g}\|_1 - \nu \|\mathbf{g}(\mathbf{w})\|_1$$

i.e.

$$\frac{\mathrm{d}}{\mathrm{d}t} T_1(\mathbf{w} + t\Delta \mathbf{w})\Big|_{t=0} \leq -\Delta \mathbf{w}^\top H \Delta \mathbf{w} + \left(\|\boldsymbol{\lambda}\|_\infty - \nu\right) \|\mathbf{g}(\mathbf{w})\|_1 < 0$$

holds if SOSC (reduced Hessian $Z^\top H Z \geq 0$) and $\|\boldsymbol{\lambda}\|_\infty > \nu$.

Outline

1. The Newton method
2. Newton on the KKT conditions
3. The reduced Newton step (unconstrained problems)
4. The merit function - Line-search for constrained problems
5. Newton-type methods
6. Sequential Quadratic Programming

Convergence of Newton-type methods

Computing $H = \nabla^2_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}, \boldsymbol{\mu})$ can be expensive, use an approximation instead !!

What happens to the convergence of the Newton method ?

Convergence of Newton-type methods

Computing $H = \nabla^2_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}, \boldsymbol{\mu})$ can be expensive, use an approximation instead !!

What happens to the convergence of the Newton method ?

Root finding problem $\mathbf{r}(\mathbf{w}) = 0$ with iteration $\mathbf{w}_{k+1} = \mathbf{w}_k - \left(\frac{\partial \mathbf{r}}{\partial \mathbf{w}}\right)^{-1} \mathbf{r}(\mathbf{w})$

Convergence of Newton-type methods

Computing $H = \nabla^2_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}, \boldsymbol{\mu})$ can be expensive, use an approximation instead !!

What happens to the convergence of the Newton method ?

Root finding problem $\mathbf{r}(\mathbf{w}) = 0$ with iteration $\mathbf{w}_{k+1} = \mathbf{w}_k - \left(\frac{\partial \mathbf{r}}{\partial \mathbf{w}}\right)^{-1} \mathbf{r}(\mathbf{w})$

The **Newton-type** iteration reads as

$$\mathbf{w}_{k+1} = \mathbf{w}_k - M_k^{-1} \mathbf{r}(\mathbf{w}_k), \qquad \text{with} \qquad M_k \text{ invertible, approximates } J = \frac{\partial \mathbf{r}(\mathbf{w})}{\partial \mathbf{w}}$$

Convergence of Newton-type methods

Computing $H = \nabla^2_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}, \boldsymbol{\mu})$ can be expensive, use an approximation instead !!

What happens to the convergence of the Newton method ?

Root finding problem $\mathbf{r}(\mathbf{w}) = 0$ with iteration $\mathbf{w}_{k+1} = \mathbf{w}_k - \left(\frac{\partial \mathbf{r}}{\partial \mathbf{w}}\right)^{-1} \mathbf{r}(\mathbf{w})$

The **Newton-type** iteration reads as

$$\mathbf{w}_{k+1} = \mathbf{w}_k - M_k^{-1} \mathbf{r}(\mathbf{w}_k), \qquad \text{with} \qquad M_k \text{ invertible, approximates } J = \frac{\partial \mathbf{r}(\mathbf{w})}{\partial \mathbf{w}}$$

We assume:

Convergence of Newton-type methods

Computing $H = \nabla^2_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}, \boldsymbol{\mu})$ can be expensive, use an approximation instead !!

What happens to the convergence of the Newton method ?

Root finding problem $\mathbf{r}(\mathbf{w}) = 0$ with iteration $\mathbf{w}_{k+1} = \mathbf{w}_k - \left(\frac{\partial \mathbf{r}}{\partial \mathbf{w}}\right)^{-1} \mathbf{r}(\mathbf{w})$

The **Newton-type** iteration reads as

$$\mathbf{w}_{k+1} = \mathbf{w}_k - M_k^{-1} \mathbf{r}(\mathbf{w}_k), \qquad \text{with} \qquad M_k \text{ invertible, approximates } J = \frac{\partial \mathbf{r}(\mathbf{w})}{\partial \mathbf{w}}$$

We assume:

- Lipschitz condition: $\|M_k^{-1}(J(\mathbf{w}_k) - J(\mathbf{w}^*))\| \leq \omega \|\mathbf{w}_k - \mathbf{w}^*\|$

Convergence of Newton-type methods

Computing $H = \nabla^2_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}, \boldsymbol{\mu})$ can be expensive, use an approximation instead !!

What happens to the convergence of the Newton method ?

Root finding problem $\mathbf{r}(\mathbf{w}) = 0$ with iteration $\mathbf{w}_{k+1} = \mathbf{w}_k - \left(\frac{\partial \mathbf{r}}{\partial \mathbf{w}}\right)^{-1} \mathbf{r}(\mathbf{w})$

The **Newton-type** iteration reads as

$$\mathbf{w}_{k+1} = \mathbf{w}_k - M_k^{-1} \mathbf{r}(\mathbf{w}_k), \qquad \text{with} \qquad M_k \text{ invertible, approximates } J = \frac{\partial \mathbf{r}(\mathbf{w})}{\partial \mathbf{w}}$$

We assume:

- Lipschitz condition: $\|M_k^{-1}(J(\mathbf{w}_k) - J(\mathbf{w}^*))\| \leq \omega \|\mathbf{w}_k - \mathbf{w}^*\|$
- Bound on the Jacobian approximation error $\|M_k^{-1}(J(\mathbf{w}_k) - M_k)\| \leq \kappa_k < \kappa < 1$

Convergence of Newton-type methods

Computing $H = \nabla^2_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}, \boldsymbol{\mu})$ can be expensive, use an approximation instead !!

What happens to the convergence of the Newton method ?

Root finding problem $\mathbf{r}(\mathbf{w}) = 0$ with iteration $\mathbf{w}_{k+1} = \mathbf{w}_k - \left(\frac{\partial \mathbf{r}}{\partial \mathbf{w}}\right)^{-1} \mathbf{r}(\mathbf{w})$

The **Newton-type** iteration reads as

$$\mathbf{w}_{k+1} = \mathbf{w}_k - M_k^{-1} \mathbf{r}(\mathbf{w}_k), \qquad \text{with} \qquad M_k \text{ invertible, approximates } J = \frac{\partial \mathbf{r}(\mathbf{w})}{\partial \mathbf{w}}$$

We assume:

- Lipschitz condition: $\|M_k^{-1}(J(\mathbf{w}_k) - J(\mathbf{w}^*))\| \leq \omega \|\mathbf{w}_k - \mathbf{w}^*\|$
- Bound on the Jacobian approximation error $\|M_k^{-1}(J(\mathbf{w}_k) - M_k)\| \leq \kappa_k < \kappa < 1$
- Good initial guess $\|\mathbf{w}_0 - \mathbf{w}^*\| \leq \frac{2}{\omega}(1 - \kappa)$

Convergence of Newton-type methods

Computing $H = \nabla^2_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}, \boldsymbol{\mu})$ can be expensive, use an approximation instead !!

What happens to the convergence of the Newton method ?

Root finding problem $\mathbf{r}(\mathbf{w}) = 0$ with iteration $\mathbf{w}_{k+1} = \mathbf{w}_k - \left(\frac{\partial \mathbf{r}}{\partial \mathbf{w}}\right)^{-1} \mathbf{r}(\mathbf{w})$

The **Newton-type** iteration reads as

$$\mathbf{w}_{k+1} = \mathbf{w}_k - M_k^{-1} \mathbf{r}(\mathbf{w}_k), \qquad \text{with} \qquad M_k \text{ invertible, approximates } J = \frac{\partial \mathbf{r}(\mathbf{w})}{\partial \mathbf{w}}$$

We assume:

- Lipschitz condition: $\|M_k^{-1}(J(\mathbf{w}_k) - J(\mathbf{w}^*))\| \leq \omega \|\mathbf{w}_k - \mathbf{w}^*\|$
- Bound on the Jacobian approximation error $\|M_k^{-1}(J(\mathbf{w}_k) - M_k)\| \leq \kappa_k < \kappa < 1$
- Good initial guess $\|\mathbf{w}_0 - \mathbf{w}^*\| \leq \frac{2}{\omega}(1 - \kappa)$

Then $\mathbf{w}_k \to \mathbf{w}^*$ with the following linear contraction in each iteration:

$$\|\mathbf{w}_{k+1} - \mathbf{w}^*\| \quad \leq \quad \left(\kappa_k + \frac{\omega}{2}\|\mathbf{w}_k - \mathbf{w}^*\|\right) \|\mathbf{w}_k - \mathbf{w}^*\|.$$

Newton-type Techniques (unconstrained)

Computing $H = \nabla^2_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}, \boldsymbol{\mu})$ can be expensive, use an approximation instead !!

Newton-type Techniques (unconstrained)

Computing $H = \nabla^2_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}, \boldsymbol{\mu})$ can be expensive, use an approximation instead !!

Descent direction

If $M_k \succ 0$ then $\Delta \mathbf{w} = -M_k^{-1} \nabla \Phi(\mathbf{w})$ is a descent direction.

Newton-type Techniques (unconstrained)

Computing $H = \nabla^2_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}, \boldsymbol{\mu})$ can be expensive, use an approximation instead !!

Descent direction

If $M_k \succ 0$ then $\Delta \mathbf{w} = -M_k^{-1} \nabla \Phi(\mathbf{w})$ is a descent direction.

Local convergence for all Newton Methods

Assume

- $\mathbf{w}^*$ is **SOSC**
- **Lipschitz Hessian**: $\| M_k^{-1} \left(\nabla^2 \Phi(\mathbf{w}_k) - \nabla^2 \Phi(\mathbf{y}) \right) \| \leq \omega \| \mathbf{w}_k - \mathbf{y} \|$ holds on the iterations, $\forall \mathbf{y}$.
- **Compatibility**: $\| M_k^{-1} \left(\nabla^2 \Phi(\mathbf{w}_k) - M_k \right) \| \leq \kappa_k$ with $\kappa_k \leq \kappa < 1$

Newton-type Techniques (unconstrained)

Computing $H = \nabla^2_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}, \boldsymbol{\mu})$ can be expensive, use an approximation instead !!

Descent direction

If $M_k \succ 0$ then $\Delta \mathbf{w} = -M_k^{-1} \nabla \Phi(\mathbf{w})$ is a descent direction.

Local convergence for all Newton Methods

Assume

- $\mathbf{w}^*$ is **SOSC**
- **Lipschitz Hessian**: $\| M_k^{-1} \left(\nabla^2 \Phi(\mathbf{w}_k) - \nabla^2 \Phi(\mathbf{y}) \right) \| \leq \omega \| \mathbf{w}_k - \mathbf{y} \|$ holds on the iterations, $\forall \mathbf{y}$.
- **Compatibility**: $\| M_k^{-1} \left(\nabla^2 \Phi(\mathbf{w}_k) - M_k \right) \| \leq \kappa_k$ with $\kappa_k \leq \kappa < 1$

Then if $\mathbf{w}_k$ is close to $\mathbf{w}^*$, $\mathbf{w}_k \to \mathbf{w}^*$ and convergence is

- Quadratic for $\kappa = 0$: $\| \mathbf{w}_{k+1} - \mathbf{w}^* \| \leq C \| \mathbf{w}_k - \mathbf{w}^* \|^2$ with $C = \omega / 2$
- Superlinear for $\kappa_k \to 0$: $\| \mathbf{w}_{k+1} - \mathbf{w}^* \| \leq C_k \| \mathbf{w}_k - \mathbf{w}^* \|$ with $C_k \to 0$
- Linear for $\kappa_k > \rho$: $\| \mathbf{w}_{k+1} - \mathbf{w}^* \| \leq C \| \mathbf{w}_k - \mathbf{w}^* \|$ with $C < 1$

Newton-type Methods (unconstrained)

Steepest descent

Use $M_k = \alpha_k^{-1} I$, then:

$$\Delta \mathbf{w}_k = -M_k^{-1} \nabla \Phi(\mathbf{w}_k) = -\alpha_k \nabla \Phi(\mathbf{w}_k)$$

Step size α_k is chosen sufficiently small by the line-search.

Convergence

- Compatibility: $\|\alpha_k \nabla^2 \Phi(\mathbf{w}_k) - I\| \leq \kappa_k$ with $\kappa_k \leq \kappa < 1$
- Constant κ_k does not converge to 0, i.e. $\kappa_k > \rho, \ \forall k$
- Linear convergence when $\mathbf{w}_k$ is close to $\mathbf{w}^*$

Newton-type Methods (unconstrained)

Cost function of the type $\Phi(\mathbf{w}) = \frac{1}{2}\|\mathbf{R}(\mathbf{w})\|^2$, with $\mathbf{R}(\mathbf{w}) \in \mathbb{R}^m$

Gauss-Newton Hessian approximation

Observe that

$$H(\mathbf{w}) = \frac{\partial}{\partial \mathbf{w}} \left(\nabla \mathbf{R}(\mathbf{w}) \mathbf{R}(\mathbf{w}) \right) = \nabla \mathbf{R}(\mathbf{w}) \nabla \mathbf{R}(\mathbf{w})^\top + \sum_{i=1}^{m} \nabla^2 \mathbf{R}_i(\mathbf{w}) \mathbf{R}_i(\mathbf{w})$$

Newton-type Methods (unconstrained)

Cost function of the type $\Phi(\mathbf{w}) = \frac{1}{2}\|\mathbf{R}(\mathbf{w})\|^2$, with $\mathbf{R}(\mathbf{w}) \in \mathbb{R}^m$

Gauss-Newton Hessian approximation

Observe that

$$H(\mathbf{w}) = \frac{\partial}{\partial \mathbf{w}} \left(\nabla \mathbf{R}(\mathbf{w}) \mathbf{R}(\mathbf{w})\right) = \nabla \mathbf{R}(\mathbf{w}) \nabla \mathbf{R}(\mathbf{w})^\top + \sum_{i=1}^{m} \nabla^2 \mathbf{R}_i(\mathbf{w}) \mathbf{R}_i(\mathbf{w})$$

Gauss-Newton method proposes to use: $\boxed{B_k = \nabla \mathbf{R}(\mathbf{w}_k) \nabla \mathbf{R}(\mathbf{w}_k)^\top + \alpha_k I}$

Newton-type Methods (unconstrained)

Cost function of the type $\Phi(\mathbf{w}) = \frac{1}{2}\|\mathbf{R}(\mathbf{w})\|^2$, with $\mathbf{R}(\mathbf{w}) \in \mathbb{R}^m$

Gauss-Newton Hessian approximation

Observe that

$$H(\mathbf{w}) = \frac{\partial}{\partial \mathbf{w}} \left(\nabla \mathbf{R}(\mathbf{w}) \mathbf{R}(\mathbf{w})\right) = \nabla \mathbf{R}(\mathbf{w}) \nabla \mathbf{R}(\mathbf{w})^\top + \sum_{i=1}^{m} \nabla^2 \mathbf{R}_i(\mathbf{w}) \mathbf{R}_i(\mathbf{w})$$

Gauss-Newton method proposes to use: $\boxed{B_k = \nabla \mathbf{R}(\mathbf{w}_k) \nabla \mathbf{R}(\mathbf{w}_k)^\top + \alpha_k I}$

B_k is a good approximation if:

- all $\nabla^2 \mathbf{R}_i(\mathbf{w})$ are small ($\mathbf{R}$ close to linear), or
- all $\mathbf{R}_i(\mathbf{w})$ are small

Newton-type Methods (unconstrained)

Cost function of the type $\Phi(\mathbf{w}) = \frac{1}{2}\|\mathbf{R}(\mathbf{w})\|^2$, with $\mathbf{R}(\mathbf{w}) \in \mathbb{R}^m$

Gauss-Newton Hessian approximation

Observe that

$$H(\mathbf{w}) = \frac{\partial}{\partial \mathbf{w}} \left(\nabla \mathbf{R}(\mathbf{w}) \mathbf{R}(\mathbf{w})\right) = \nabla \mathbf{R}(\mathbf{w}) \nabla \mathbf{R}(\mathbf{w})^\top + \sum_{i=1}^{m} \nabla^2 \mathbf{R}_i(\mathbf{w}) \mathbf{R}_i(\mathbf{w})$$

Gauss-Newton method proposes to use: $B_k = \nabla \mathbf{R}(\mathbf{w}_k) \nabla \mathbf{R}(\mathbf{w}_k)^\top + \alpha_k I$

B_k is a good approximation if:

- all $\nabla^2 \mathbf{R}_i(\mathbf{w})$ are small ($\mathbf{R}$ close to linear), or
- all $\mathbf{R}_i(\mathbf{w})$ are small

Typical application to fitting problems or control problems: $\mathbf{R}(\mathbf{w}) = \mathbf{y}(\mathbf{w}) - \bar{\mathbf{y}}$

Newton-type Methods (unconstrained)

Cost function of the type $\Phi(\mathbf{w}) = \frac{1}{2}\|\mathbf{R}(\mathbf{w})\|^2$, with $\mathbf{R}(\mathbf{w}) \in \mathbb{R}^m$

Gauss-Newton Hessian approximation

Observe that

$$H(\mathbf{w}) = \frac{\partial}{\partial \mathbf{w}} \left(\nabla \mathbf{R}(\mathbf{w}) \mathbf{R}(\mathbf{w})\right) = \nabla \mathbf{R}(\mathbf{w}) \nabla \mathbf{R}(\mathbf{w})^\top + \sum_{i=1}^{m} \nabla^2 \mathbf{R}_i(\mathbf{w}) \mathbf{R}_i(\mathbf{w})$$

Gauss-Newton method proposes to use: $B_k = \nabla \mathbf{R}(\mathbf{w}_k) \nabla \mathbf{R}(\mathbf{w}_k)^\top + \alpha_k I$

B_k is a good approximation if:

- all $\nabla^2 \mathbf{R}_i(\mathbf{w})$ are small ($\mathbf{R}$ close to linear), or
- all $\mathbf{R}_i(\mathbf{w})$ are small

Typical application to fitting problems or control problems: $\mathbf{R}(\mathbf{w}) = \mathbf{y}(\mathbf{w}) - \bar{\mathbf{y}}$

Convergence

- If $\sum_{i=1}^{m} \nabla^2 \mathbf{R}_i(\mathbf{w}) \mathbf{R}_i(\mathbf{w}) \to 0$ then $\kappa_k \to 0$
- Superlinear convergence in some cases...

Newton-type Methods (unconstrained)

Compute $H(\mathbf{w})$ numerically in an efficient (iterative) way

BFGS

Define

$$\begin{aligned} \mathbf{s}_k &= \mathbf{w}_{k+1} - \mathbf{w}_k \\ \mathbf{y}_k &= \nabla\Phi(\mathbf{w}_{k+1}) - \nabla\Phi(\mathbf{w}_k) \end{aligned}$$

Idea: Update $M_k \rightarrow M_{k+1}$ such that $M_{k+1}\mathbf{s}_k = \mathbf{y}_k$ (secant condition)

Newton-type Methods (unconstrained)

Compute $H(\mathbf{w})$ numerically in an efficient (iterative) way

BFGS

Define

$$\begin{aligned} \mathbf{s}_k &= \mathbf{w}_{k+1} - \mathbf{w}_k \\ \mathbf{y}_k &= \nabla\Phi(\mathbf{w}_{k+1}) - \nabla\Phi(\mathbf{w}_k) \end{aligned}$$

Idea: Update $M_k \rightarrow M_{k+1}$ such that $M_{k+1}\mathbf{s}_k = \mathbf{y}_k$ (secant condition)

BFGS formula:

$$M_{k+1} = M_k - \frac{M_k \mathbf{s}\mathbf{s}^\top M_k}{\mathbf{s}^\top M_k \mathbf{s}} + \frac{\mathbf{y}\mathbf{y}^\top}{\mathbf{s}^\top \mathbf{y}}, \quad M_0 = I$$

Newton-type Methods (unconstrained)

Compute $H(\mathbf{w})$ numerically in an efficient (iterative) way

BFGS

Define

$$\begin{aligned} \mathbf{s}_k &= \mathbf{w}_{k+1} - \mathbf{w}_k \\ \mathbf{y}_k &= \nabla\Phi(\mathbf{w}_{k+1}) - \nabla\Phi(\mathbf{w}_k) \end{aligned}$$

Idea: Update $M_k \rightarrow M_{k+1}$ such that $M_{k+1}\mathbf{s}_k = \mathbf{y}_k$ (secant condition)

BFGS formula: $$M_{k+1} = M_k - \frac{M_k \mathbf{s}\mathbf{s}^\top M_k}{\mathbf{s}^\top M_k \mathbf{s}} + \frac{\mathbf{y}\mathbf{y}^\top}{\mathbf{s}^\top \mathbf{y}}, \quad M_0 = I$$

See "Powell's trick" ensure $M_{k+1} > 0$

Newton-type Methods (unconstrained)

Compute $H(\mathbf{w})$ numerically in an efficient (iterative) way

BFGS

Define

$$\begin{aligned} \mathbf{s}_k &= \mathbf{w}_{k+1} - \mathbf{w}_k \\ \mathbf{y}_k &= \nabla\Phi(\mathbf{w}_{k+1}) - \nabla\Phi(\mathbf{w}_k) \end{aligned}$$

Idea: Update $M_k \rightarrow M_{k+1}$ such that $M_{k+1}\mathbf{s}_k = \mathbf{y}_k$ (secant condition)

BFGS formula: $$M_{k+1} = M_k - \frac{M_k \mathbf{s}\mathbf{s}^\top M_k}{\mathbf{s}^\top M_k \mathbf{s}} + \frac{\mathbf{y}\mathbf{y}^\top}{\mathbf{s}^\top \mathbf{y}}, \quad M_0 = I$$

See "Powell's trick" ensure $M_{k+1} > 0$

Convergence

- It can be shown that $M_k \rightarrow H(\mathbf{w})$, then $\kappa_k \rightarrow 0$
- Superlinear convergence after enough (and "sufficiently rich") iterations

Newton-type Methods (unconstrained)

Compute $H(\mathbf{w})$ numerically in an efficient (iterative) way

BFGS

Define

$$\begin{aligned} \mathbf{s}_k &= \mathbf{w}_{k+1} - \mathbf{w}_k \\ \mathbf{y}_k &= \nabla\Phi(\mathbf{w}_{k+1}) - \nabla\Phi(\mathbf{w}_k) \end{aligned}$$

Idea: Update $M_k \to M_{k+1}$ such that $M_{k+1}\mathbf{s}_k = \mathbf{y}_k$ (secant condition)

BFGS formula: $$M_{k+1} = M_k - \frac{M_k \mathbf{s}\mathbf{s}^\top M_k}{\mathbf{s}^\top M_k \mathbf{s}} + \frac{\mathbf{y}\mathbf{y}^\top}{\mathbf{s}^\top \mathbf{y}}, \quad M_0 = I$$

See "Powell's trick" ensure $M_{k+1} > 0$

Convergence

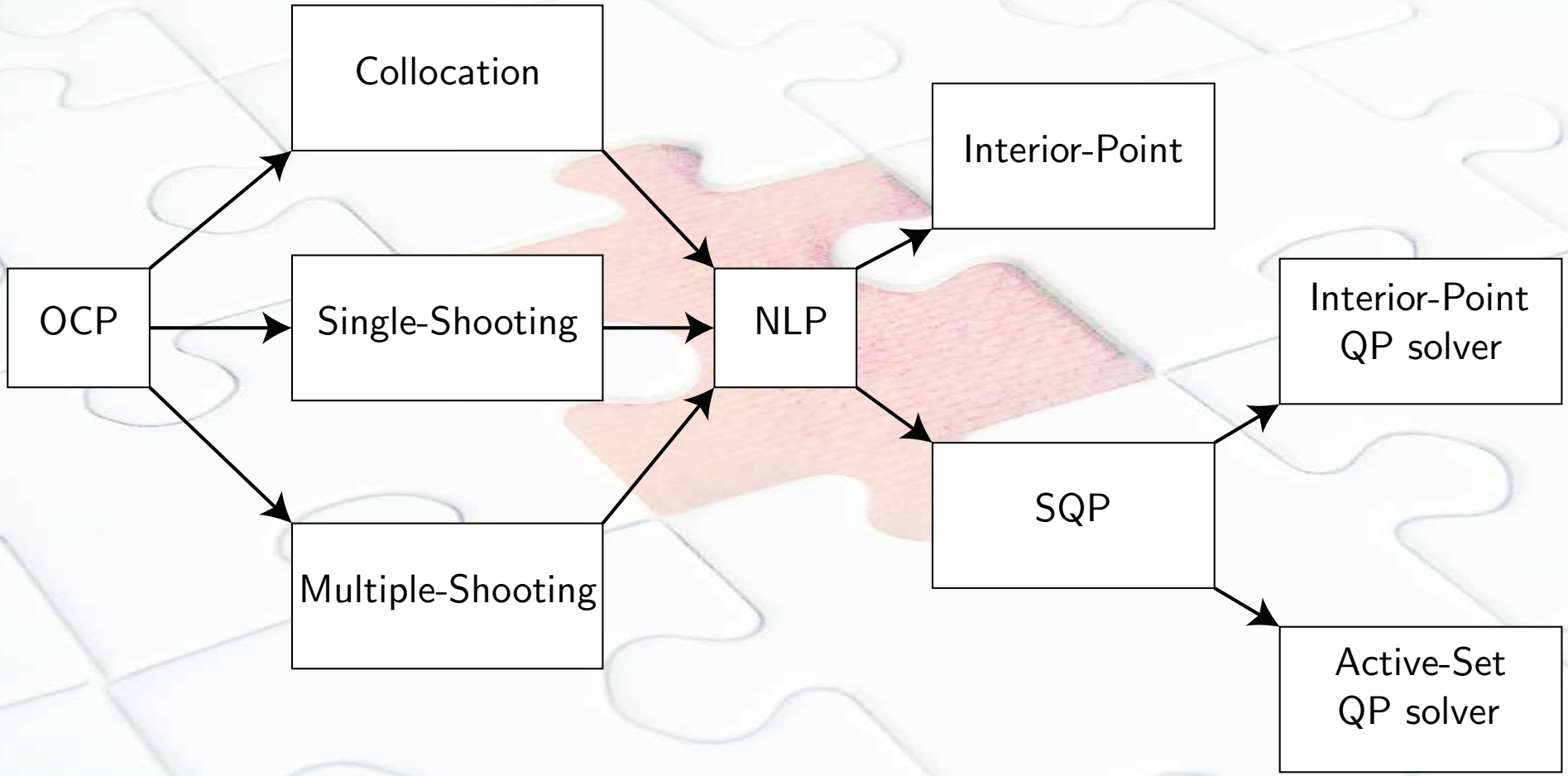
- It can be shown that $M_k \to H(\mathbf{w})$, then $\kappa_k \to 0$
- Superlinear convergence after enough (and "sufficiently rich") iterations

Obs: BFGS typically yields **dense Hessian** approximations, whereas the exact Hessian may be sparse. Factorization can become *unnecessarily* expensive !!

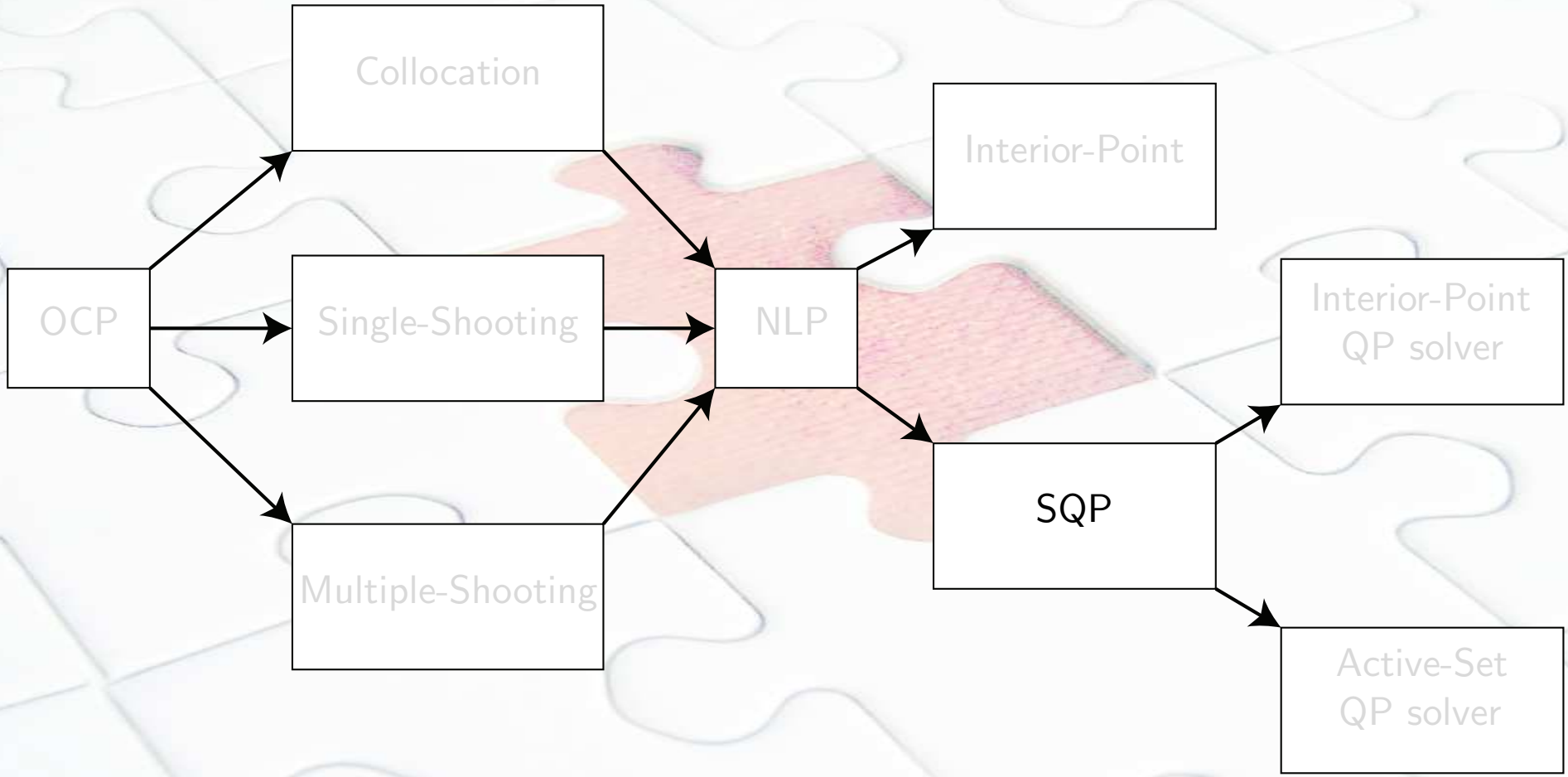
Outline

1. The Newton method
2. Newton on the KKT conditions
3. The reduced Newton step (unconstrained problems)
4. The merit function - Line-search for constrained problems
5. Newton-type methods
6. Sequential Quadratic Programming

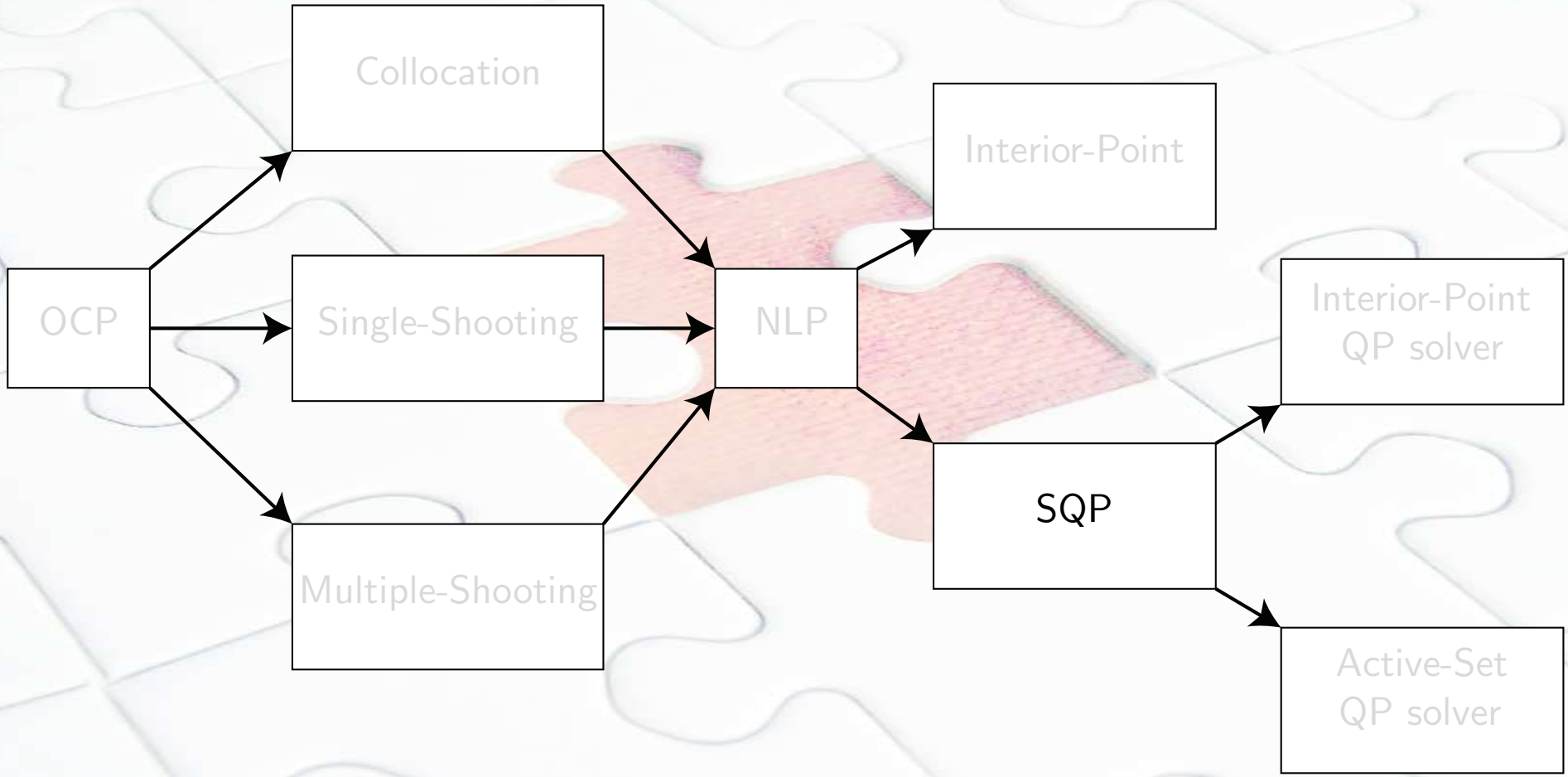
Survival map of Direct Optimal Control



Survival map of Direct Optimal Control



Survival map of Direct Optimal Control



What about inequality constraints ?

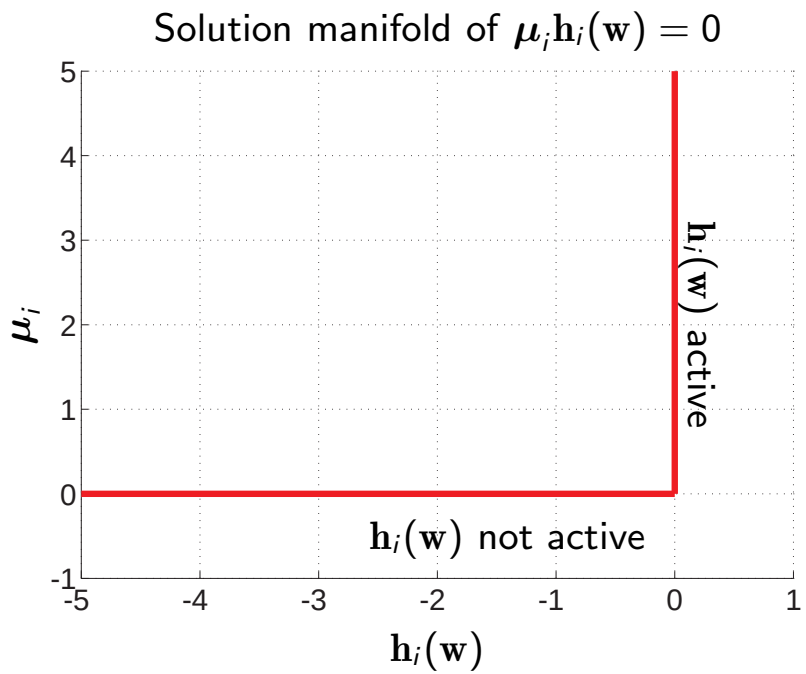
Find the "primal-dual" variables $\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\lambda}$ such that:

Primal Feasibility:	$\mathbf{g}(\mathbf{w}) = 0, \quad \mathbf{h}(\mathbf{w}) \leq 0,$
Dual Feasibility:	$\nabla_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \boldsymbol{\mu}, \boldsymbol{\lambda}) = 0, \quad \boldsymbol{\mu} \geq 0,$
Complementarity Slackness:	$\mu_i \mathbf{h}_i(\mathbf{w}) = 0, \quad i = 1, \ldots$

What about inequality constraints ?

Find the "primal-dual" variables $\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\lambda}$ such that:

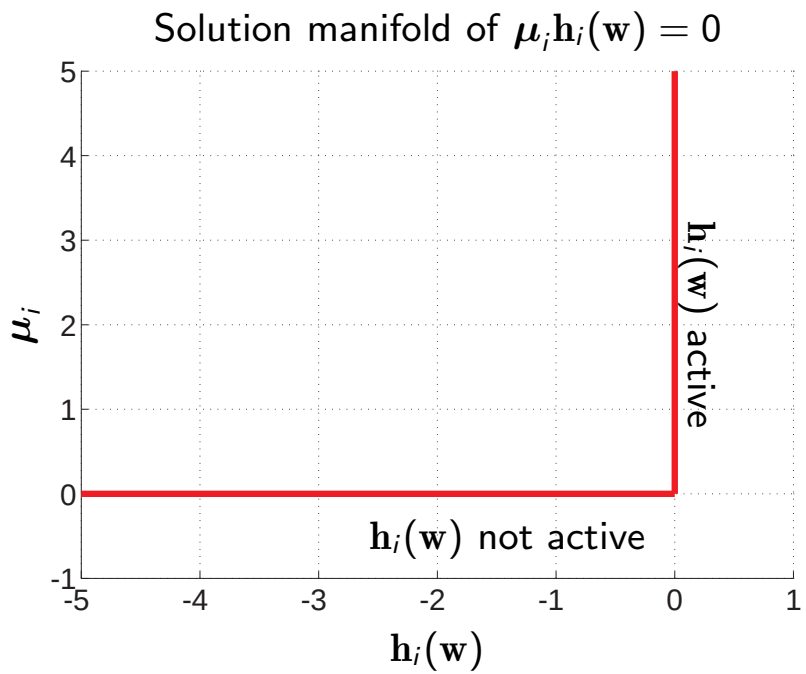
Primal Feasibility:	$\mathbf{g}(\mathbf{w}) = 0, \quad \mathbf{h}(\mathbf{w}) \leq 0,$
Dual Feasibility:	$\nabla_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \boldsymbol{\mu}, \boldsymbol{\lambda}) = 0, \quad \boldsymbol{\mu} \geq 0,$
Complementarity Slackness:	$\mu_i \mathbf{h}_i(\mathbf{w}) = 0, \quad i = 1, \ldots$



What about inequality constraints ?

Find the "primal-dual" variables $\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\lambda}$ such that:

Primal Feasibility:	$\mathbf{g}(\mathbf{w}) = 0, \quad \mathbf{h}(\mathbf{w}) \leq 0,$
Dual Feasibility:	$\nabla_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \boldsymbol{\mu}, \boldsymbol{\lambda}) = 0, \quad \boldsymbol{\mu} \geq 0,$
Complementarity Slackness:	$\mu_i \mathbf{h}_i(\mathbf{w}) = 0, \quad i = 1, \ldots$



Manifold generated by the Complementarity Slackness condition is not-smooth, Newton can not be used !!

Quadratic model interpretation

Problem:

$$\begin{array}{ll} \min\limits_{\mathbf{w}} & \Phi(\mathbf{w}) \\ \text{s.t.} & \mathbf{g}(\mathbf{w}) = 0 \end{array}$$

The Newton direction is given by

$$\begin{bmatrix} H(\mathbf{w},\boldsymbol{\lambda}) & \nabla \mathbf{g}(\mathbf{w}) \\ \nabla \mathbf{g}(\mathbf{w})^{\mathsf{T}} & 0 \end{bmatrix} \begin{bmatrix} \Delta \mathbf{w} \\ \boldsymbol{\lambda}^{+} \end{bmatrix} = - \begin{bmatrix} \nabla \Phi(\mathbf{w}) \\ \mathbf{g}(\mathbf{w}) \end{bmatrix}$$

with $H(\mathbf{w},\boldsymbol{\lambda}) = \nabla^2_{\mathbf{w}} \mathcal{L}(\mathbf{w},\boldsymbol{\lambda})$

Quadratic model interpretation

Problem:

$$\begin{array}{ll} \min\limits_{\mathbf{w}} & \Phi(\mathbf{w}) \\ \text{s.t.} & \mathbf{g}(\mathbf{w}) = 0 \end{array}$$

The Newton direction is given by

$$\begin{bmatrix} H(\mathbf{w},\boldsymbol{\lambda}) & \nabla \mathbf{g}(\mathbf{w}) \\ \nabla \mathbf{g}(\mathbf{w})^{\mathsf{T}} & 0 \end{bmatrix} \begin{bmatrix} \Delta \mathbf{w} \\ \boldsymbol{\lambda}^{+} \end{bmatrix} = - \begin{bmatrix} \nabla \Phi(\mathbf{w}) \\ \mathbf{g}(\mathbf{w}) \end{bmatrix}$$

with $H(\mathbf{w},\boldsymbol{\lambda}) = \nabla^2_{\mathbf{w}} \mathcal{L}(\mathbf{w},\boldsymbol{\lambda})$

The Newton direction is given by the Quadratic Program (QP):

$$\begin{array}{ll} \min\limits_{\Delta \mathbf{w}} & \frac{1}{2} \Delta \mathbf{w}^{\mathsf{T}} H(\mathbf{w},\boldsymbol{\lambda}) \Delta \mathbf{w} + \nabla \Phi(\mathbf{w})^{\mathsf{T}} \Delta \mathbf{w} \\ \text{s.t.} & \mathbf{g}(\mathbf{w}) + \nabla \mathbf{g}(\mathbf{w})^{\mathsf{T}} \Delta \mathbf{w} = 0 \end{array}$$

Quadratic model interpretation

Problem:

$$\begin{array}{ll} \min\limits_{\mathbf{w}} & \Phi(\mathbf{w}) \\ \text{s.t.} & \mathbf{g}(\mathbf{w}) = 0 \end{array}$$

The Newton direction is given by

$$\begin{bmatrix} H(\mathbf{w},\boldsymbol{\lambda}) & \nabla \mathbf{g}(\mathbf{w}) \\ \nabla \mathbf{g}(\mathbf{w})^{\mathsf{T}} & 0 \end{bmatrix} \begin{bmatrix} \Delta \mathbf{w} \\ \boldsymbol{\lambda}^{+} \end{bmatrix} = - \begin{bmatrix} \nabla \Phi(\mathbf{w}) \\ \mathbf{g}(\mathbf{w}) \end{bmatrix}$$

with $H(\mathbf{w},\boldsymbol{\lambda}) = \nabla^2_{\mathbf{w}} \mathcal{L}(\mathbf{w},\boldsymbol{\lambda})$

The Newton direction is given by the Quadratic Program (QP):

$$\begin{array}{ll} \min\limits_{\Delta \mathbf{w}} & \frac{1}{2} \Delta \mathbf{w}^{\mathsf{T}} H(\mathbf{w},\boldsymbol{\lambda}) \Delta \mathbf{w} + \nabla \Phi(\mathbf{w})^{\mathsf{T}} \Delta \mathbf{w} \\ \text{s.t.} & \mathbf{g}(\mathbf{w}) + \nabla \mathbf{g}(\mathbf{w})^{\mathsf{T}} \Delta \mathbf{w} = 0 \end{array}$$

Dual variables $\boldsymbol{\lambda}^{+}$ given by the dual variables of the QP, i.e. $\boldsymbol{\lambda}^{+} = \boldsymbol{\lambda}_{\mathrm{QP}}$

Quadratic model interpretation

Problem:

$$\begin{array}{ll} \min\limits_{\mathbf{w}} & \Phi(\mathbf{w}) \\ \text{s.t.} & \mathbf{g}(\mathbf{w}) = 0 \end{array}$$

The Newton direction is given by

$$\begin{bmatrix} H(\mathbf{w},\boldsymbol{\lambda}) & \nabla \mathbf{g}(\mathbf{w}) \\ \nabla \mathbf{g}(\mathbf{w})^{\mathsf{T}} & 0 \end{bmatrix} \begin{bmatrix} \Delta \mathbf{w} \\ \boldsymbol{\lambda}^{+} \end{bmatrix} = - \begin{bmatrix} \nabla \Phi(\mathbf{w}) \\ \mathbf{g}(\mathbf{w}) \end{bmatrix}$$

with $H(\mathbf{w},\boldsymbol{\lambda}) = \nabla^2_{\mathbf{w}} \mathcal{L}(\mathbf{w},\boldsymbol{\lambda})$

The Newton direction is given by the Quadratic Program (QP):

$$\begin{array}{ll} \min\limits_{\Delta \mathbf{w}} & \frac{1}{2} \Delta \mathbf{w}^{\mathsf{T}} H(\mathbf{w},\boldsymbol{\lambda}) \Delta \mathbf{w} + \nabla \Phi(\mathbf{w})^{\mathsf{T}} \Delta \mathbf{w} \\ \text{s.t.} & \mathbf{g}(\mathbf{w}) + \nabla \mathbf{g}(\mathbf{w})^{\mathsf{T}} \Delta \mathbf{w} = 0 \end{array}$$

Dual variables $\boldsymbol{\lambda}^{+}$ given by the dual variables of the QP, i.e. $\boldsymbol{\lambda}^{+} = \boldsymbol{\lambda}_{\mathrm{QP}}$

Proof: KKT of the QP are equivalent to the linearised KKT system of the original problem.

Quadratic interpretation for inequality constraints

Problem:

$$\begin{array}{ll} \min\limits_{\mathbf{w}} & \Phi(\mathbf{w}) \\ \text{s.t.} & \mathbf{g}(\mathbf{w}) = 0 \\ \text{s.t.} & \mathbf{h}(\mathbf{w}) \leq 0 \end{array}$$

Quadratic interpretation for inequality constraints

Problem:

$$\begin{array}{ll} \min\limits_{\mathbf{w}} & \Phi(\mathbf{w}) \\ \text{s.t.} & \mathbf{g}(\mathbf{w}) = 0 \\ \text{s.t.} & \mathbf{h}(\mathbf{w}) \leq 0 \end{array}$$

The Newton direction is given by the Quadratic Program (QP):

$$\begin{array}{ll} \min\limits_{\Delta\mathbf{w}} & \frac{1}{2}\Delta\mathbf{w}^{\mathsf{T}} H(\mathbf{w},\boldsymbol{\lambda},\boldsymbol{\mu})\Delta\mathbf{w} + \nabla\Phi(\mathbf{w})^{\mathsf{T}}\Delta\mathbf{w} \\ \text{s.t.} & \mathbf{g}(\mathbf{w}) + \nabla\mathbf{g}(\mathbf{w})^{\mathsf{T}}\Delta\mathbf{w} = 0 \\ & \mathbf{h}(\mathbf{w}) + \nabla\mathbf{h}(\mathbf{w})^{\mathsf{T}}\Delta\mathbf{w} \leq 0 \end{array}$$

with $H(\mathbf{w},\boldsymbol{\lambda}) = \nabla^2_{\mathbf{w}}\mathcal{L}(\mathbf{w},\boldsymbol{\lambda})$

Quadratic interpretation for inequality constraints

Problem:

$$\begin{array}{ll} \min\limits_{\mathbf{w}} & \Phi(\mathbf{w}) \\ \text{s.t.} & \mathbf{g}(\mathbf{w}) = 0 \\ \text{s.t.} & \mathbf{h}(\mathbf{w}) \leq 0 \end{array}$$

The Newton direction is given by the Quadratic Program (QP):

$$\begin{array}{ll} \min\limits_{\Delta\mathbf{w}} & \frac{1}{2}\Delta\mathbf{w}^{\mathsf{T}} H(\mathbf{w},\boldsymbol{\lambda},\boldsymbol{\mu})\Delta\mathbf{w} + \nabla\Phi(\mathbf{w})^{\mathsf{T}}\Delta\mathbf{w} \\ \text{s.t.} & \mathbf{g}(\mathbf{w}) + \nabla\mathbf{g}(\mathbf{w})^{\mathsf{T}}\Delta\mathbf{w} = 0 \\ & \mathbf{h}(\mathbf{w}) + \nabla\mathbf{h}(\mathbf{w})^{\mathsf{T}}\Delta\mathbf{w} \leq 0 \end{array}$$

with $H(\mathbf{w},\boldsymbol{\lambda}) = \nabla^2_{\mathbf{w}}\mathcal{L}(\mathbf{w},\boldsymbol{\lambda})$

Dual variables $\boldsymbol{\lambda}^+$ and $\boldsymbol{\mu}^+$ given by the dual variables of the QP, i.e.

$$\boldsymbol{\lambda}^+ = \boldsymbol{\lambda}_{\text{QP}}, \quad \boldsymbol{\mu}^+ = \boldsymbol{\mu}_{\text{QP}}$$

SQP - Monitoring progress with the T_1 merit function:

$$T_1(\mathbf{w}) = \Phi(\mathbf{w}) + \nu\|\mathbf{g}(\mathbf{w})\|_1 + \nu\sum_{i=1}^{m}\max(0, \mathbf{h}_i(\mathbf{w}))$$

Algorithm: SQP with line-search

Input: guess $\mathbf{w}$, $\boldsymbol{\lambda}$, $\boldsymbol{\mu}$

while $\|\nabla\mathcal{L}\|$ *or* $\|\mathbf{g}\|$ *or* $\max(0, \mathbf{h}) \geq \text{tol}$ **do**

Compute $\mathbf{g}$, $\mathbf{h}$, $\nabla\Phi(\mathbf{w})$, $\nabla\mathbf{g}(\mathbf{w})$, $\nabla\mathbf{h}(\mathbf{w})$, $H(\mathbf{w}, \boldsymbol{\mu}, \boldsymbol{\lambda})$

Compute Newton direction by solving the QP

$$\begin{aligned} \min_{\Delta\mathbf{w}} \quad & \frac{1}{2}\Delta\mathbf{w}^T H(\mathbf{w}, \boldsymbol{\lambda}, \boldsymbol{\mu})\Delta\mathbf{w} + \nabla\Phi(\mathbf{w})^T\Delta\mathbf{w} \\ \text{s.t.} \quad & \mathbf{g}(\mathbf{w}) + \nabla\mathbf{g}(\mathbf{w})^T\Delta\mathbf{w} = 0 \\ & \mathbf{h}(\mathbf{w}) + \nabla\mathbf{h}(\mathbf{w})^T\Delta\mathbf{w} \leq 0 \end{aligned}$$

Perform line-search on $T_1(\mathbf{w} + \alpha\Delta\mathbf{w})$, get step length α

Take primal step: $\mathbf{w} \leftarrow \mathbf{w} + \alpha\Delta\mathbf{w}$

Take dual step: $\boldsymbol{\lambda} \leftarrow (1-\alpha)\boldsymbol{\lambda} + \alpha\boldsymbol{\lambda}_{QP}$, $\boldsymbol{\mu} \leftarrow (1-\alpha)\boldsymbol{\mu} + \alpha\boldsymbol{\mu}_{QP}$

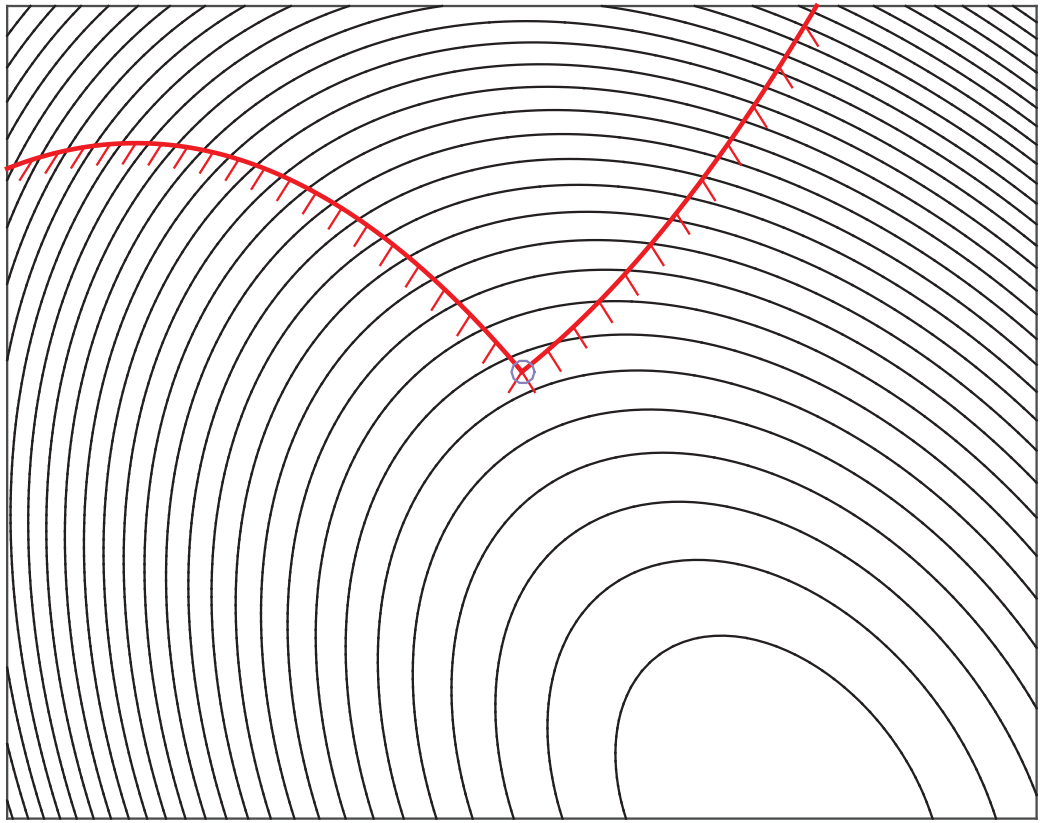
return $\mathbf{w}$, $\boldsymbol{\lambda}$, $\boldsymbol{\mu}$

If $H(\mathbf{x}_k, \boldsymbol{\mu}_k, \boldsymbol{\lambda}_k) \succ 0$ and $\nu > \max\left\{\|\boldsymbol{\mu}_{k+1}\|_\infty, \|\boldsymbol{\lambda}_{k+1}\|_\infty\right\}$ then $\Delta\mathbf{w}$ is a descent direction for $T_1(\mathbf{w})$

SQP - Illustration

NLP:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2} \left\| \mathbf{w} - \mathbf{w}_0 \right\|_Q^2 \\ \text{s.t.} \quad & \mathbf{h}\left(\mathbf{w}\right) \leq 0 \end{aligned}$$



QP:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2} \Delta \mathbf{w}^\top H\left(\mathbf{w}, \boldsymbol{\mu}\right) \Delta \mathbf{w} + \nabla \Phi\left(\mathbf{w}\right)^\top \Delta \mathbf{w} \\ \text{s.t.} \quad & \mathbf{h}\left(\mathbf{w}\right) + \nabla \mathbf{h}\left(\mathbf{w}\right)^\top \Delta \mathbf{w} \leq 0 \end{aligned}$$

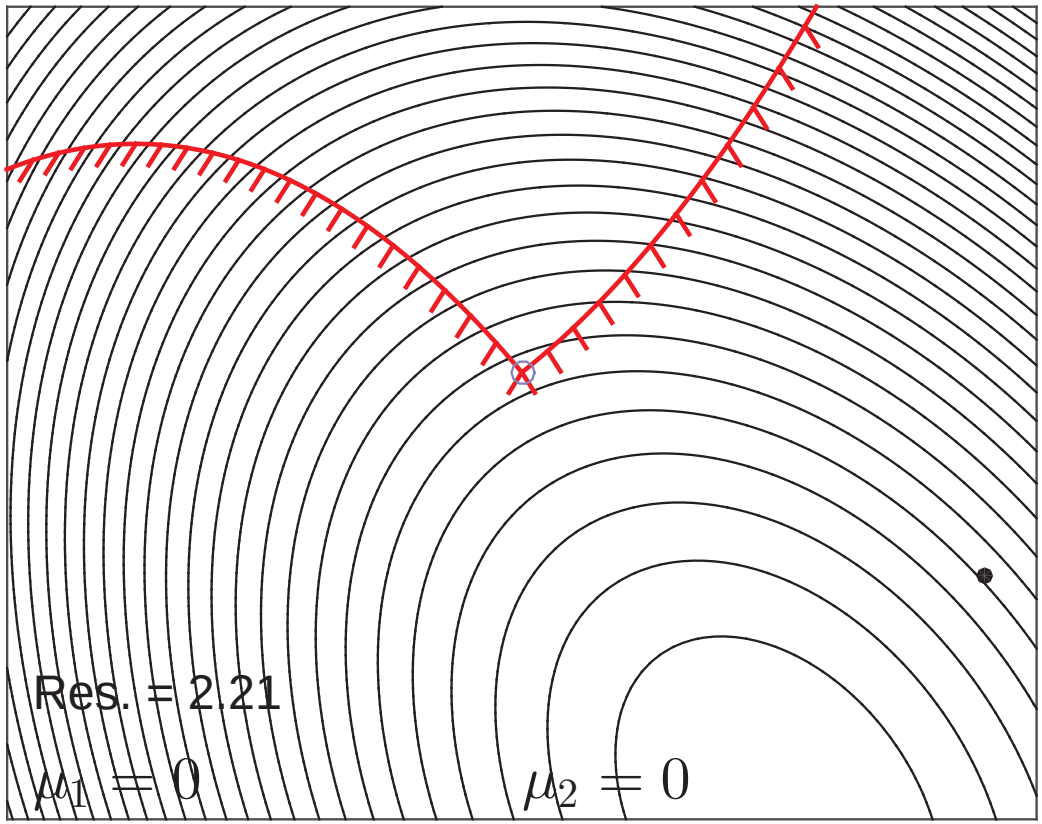
Hessian:

$$H\left(\mathbf{w}, \boldsymbol{\mu}\right) = \nabla^2 \Phi\left(\mathbf{w}\right) + \nabla^2 \left(\boldsymbol{\mu}^\top \mathbf{h}\left(\mathbf{w}\right) \right)$$

SQP - Illustration

NLP:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2} \left\| \mathbf{w} - \mathbf{w}_0 \right\|_Q^2 \\ \text{s.t.} \quad & \mathbf{h}(\mathbf{w}) \leq 0 \end{aligned}$$



QP:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2} \Delta \mathbf{w}^\top H(\mathbf{w}, \boldsymbol{\mu}) \Delta \mathbf{w} + \nabla \Phi(\mathbf{w})^\top \Delta \mathbf{w} \\ \text{s.t.} \quad & \mathbf{h}(\mathbf{w}) + \nabla \mathbf{h}(\mathbf{w})^\top \Delta \mathbf{w} \leq 0 \end{aligned}$$

Hessian:

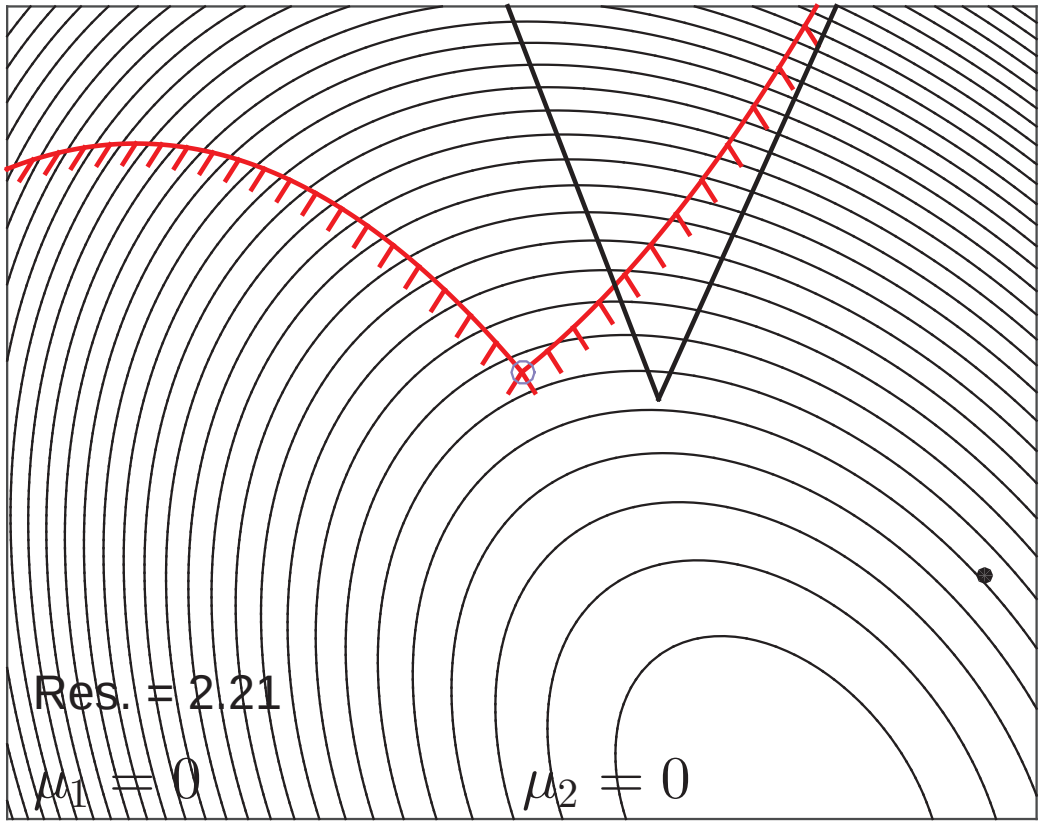
$$H(\mathbf{w}, \boldsymbol{\mu}) = \nabla^2 \Phi(\mathbf{w}) + \nabla^2 \left(\boldsymbol{\mu}^\top \mathbf{h}(\mathbf{w}) \right)$$

SQP - Illustration

NLP:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2} \left\| \mathbf{w} - \mathbf{w}_0 \right\|_Q^2 \\ \text{s.t.} \quad & \mathbf{h}(\mathbf{w}) \leq 0 \end{aligned}$$

Linearized constraints



QP:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2} \Delta \mathbf{w}^\top H(\mathbf{w}, \boldsymbol{\mu}) \Delta \mathbf{w} + \nabla \Phi(\mathbf{w})^\top \Delta \mathbf{w} \\ \text{s.t.} \quad & \mathbf{h}(\mathbf{w}) + \nabla \mathbf{h}(\mathbf{w})^\top \Delta \mathbf{w} \leq 0 \end{aligned}$$

Hessian:

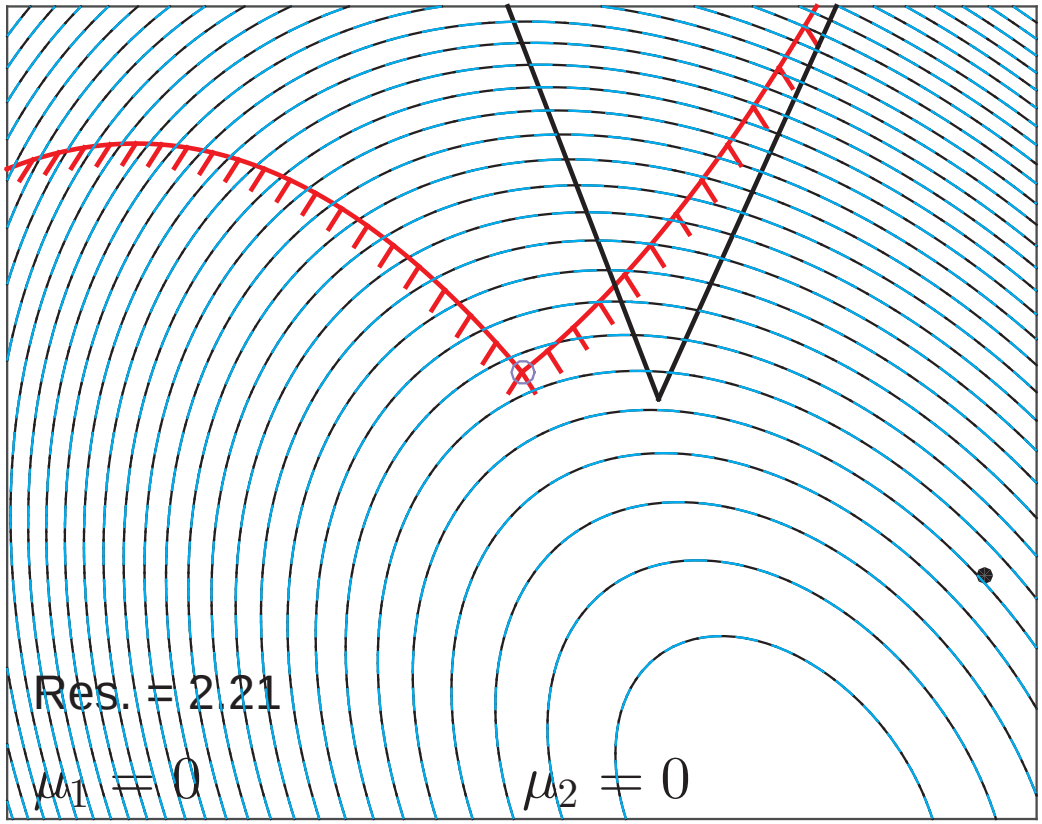
$$H(\mathbf{w}, \boldsymbol{\mu}) = \nabla^2 \Phi(\mathbf{w}) + \nabla^2 \left(\boldsymbol{\mu}^\top \mathbf{h}(\mathbf{w}) \right)$$

SQP - Illustration

NLP:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2} \left\| \mathbf{w} - \mathbf{w}_0 \right\|_Q^2 \\ \text{s.t.} \quad & \mathbf{h}(\mathbf{w}) \leq 0 \end{aligned}$$

Contours of QP cost



QP:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2} \Delta \mathbf{w}^\top H(\mathbf{w}, \boldsymbol{\mu}) \Delta \mathbf{w} + \nabla \Phi(\mathbf{w})^\top \Delta \mathbf{w} \\ \text{s.t.} \quad & \mathbf{h}(\mathbf{w}) + \nabla \mathbf{h}(\mathbf{w})^\top \Delta \mathbf{w} \leq 0 \end{aligned}$$

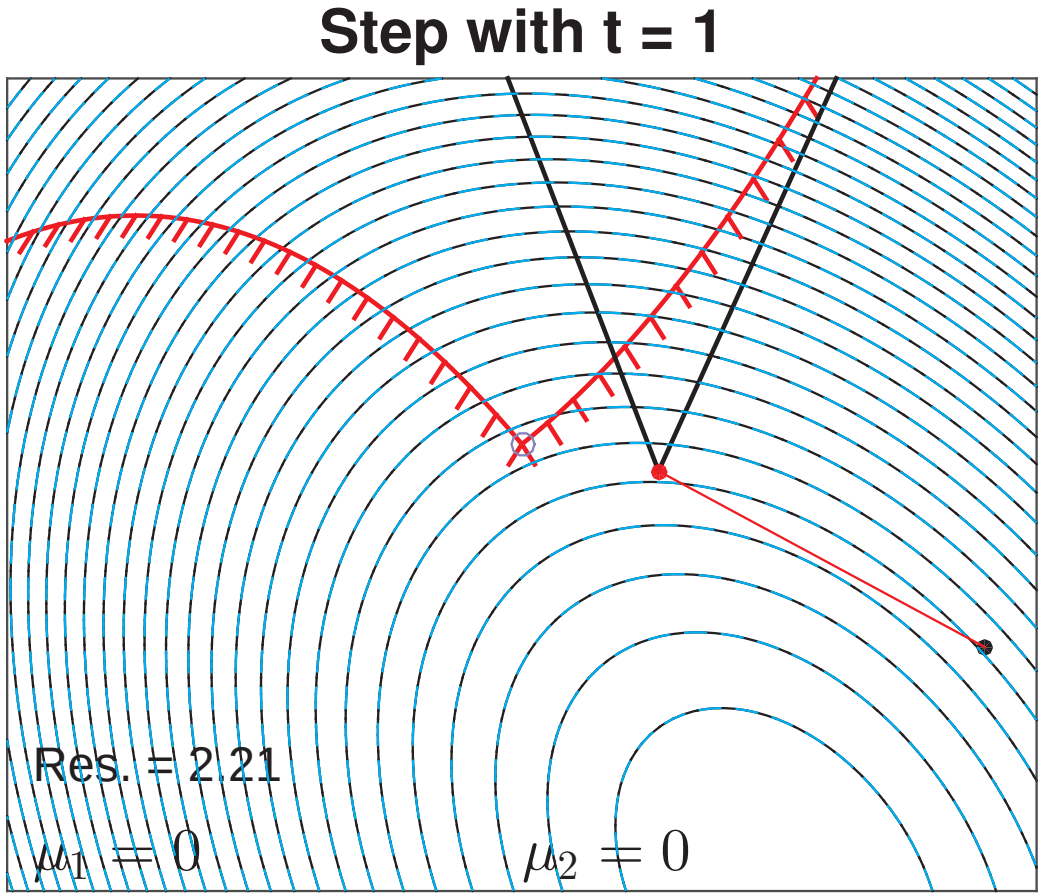
Hessian:

$$H(\mathbf{w}, \boldsymbol{\mu}) = \nabla^2 \Phi(\mathbf{w}) + \nabla^2 \left(\boldsymbol{\mu}^\top \mathbf{h}(\mathbf{w}) \right)$$

SQP - Illustration

NLP:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2} \left\| \mathbf{w} - \mathbf{w}_0 \right\|_Q^2 \\ \text{s.t.} \quad & \mathbf{h}(\mathbf{w}) \leq 0 \end{aligned}$$



QP:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2} \Delta \mathbf{w}^\top H(\mathbf{w}, \boldsymbol{\mu}) \Delta \mathbf{w} + \nabla \Phi(\mathbf{w})^\top \Delta \mathbf{w} \\ \text{s.t.} \quad & \mathbf{h}(\mathbf{w}) + \nabla \mathbf{h}(\mathbf{w})^\top \Delta \mathbf{w} \leq 0 \end{aligned}$$

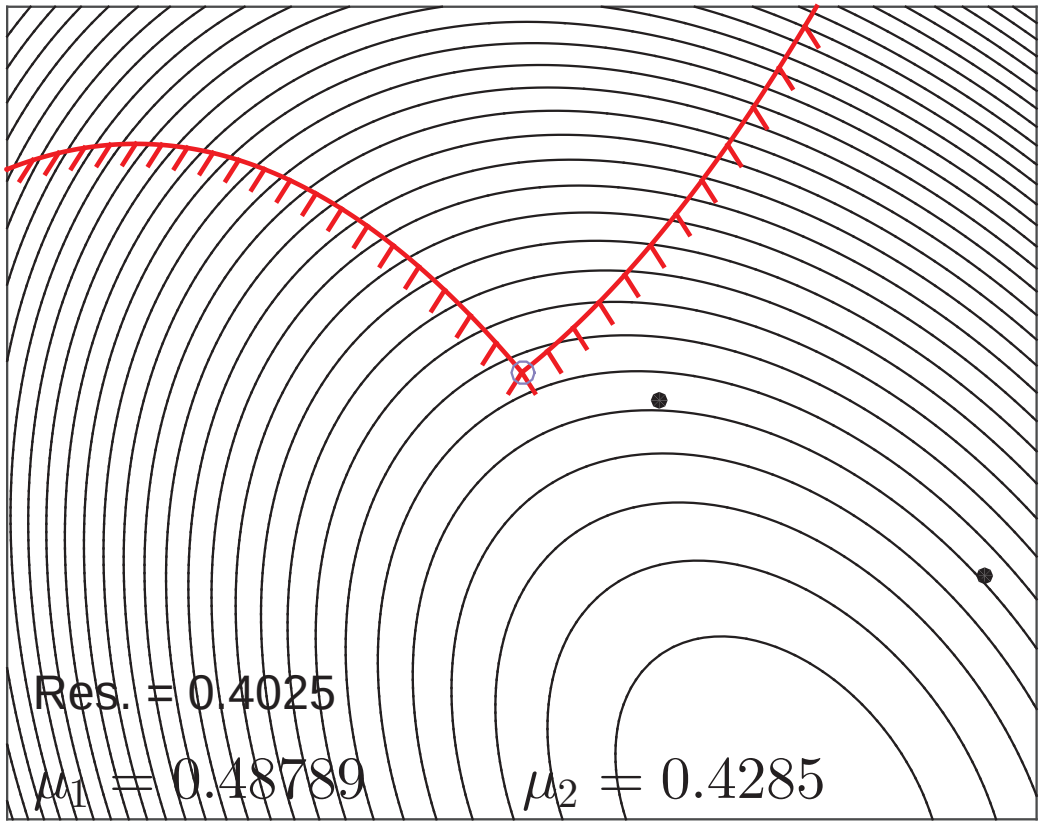
Hessian:

$$H(\mathbf{w}, \boldsymbol{\mu}) = \nabla^2 \Phi(\mathbf{w}) + \nabla^2 \left(\boldsymbol{\mu}^\top \mathbf{h}(\mathbf{w}) \right)$$

SQP - Illustration

NLP:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2} \left\| \mathbf{w} - \mathbf{w}_0 \right\|_Q^2 \\ \text{s.t.} \quad & \mathbf{h}(\mathbf{w}) \leq 0 \end{aligned}$$



QP:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2} \Delta \mathbf{w}^\top H(\mathbf{w}, \boldsymbol{\mu}) \Delta \mathbf{w} + \nabla \Phi(\mathbf{w})^\top \Delta \mathbf{w} \\ \text{s.t.} \quad & \mathbf{h}(\mathbf{w}) + \nabla \mathbf{h}(\mathbf{w})^\top \Delta \mathbf{w} \leq 0 \end{aligned}$$

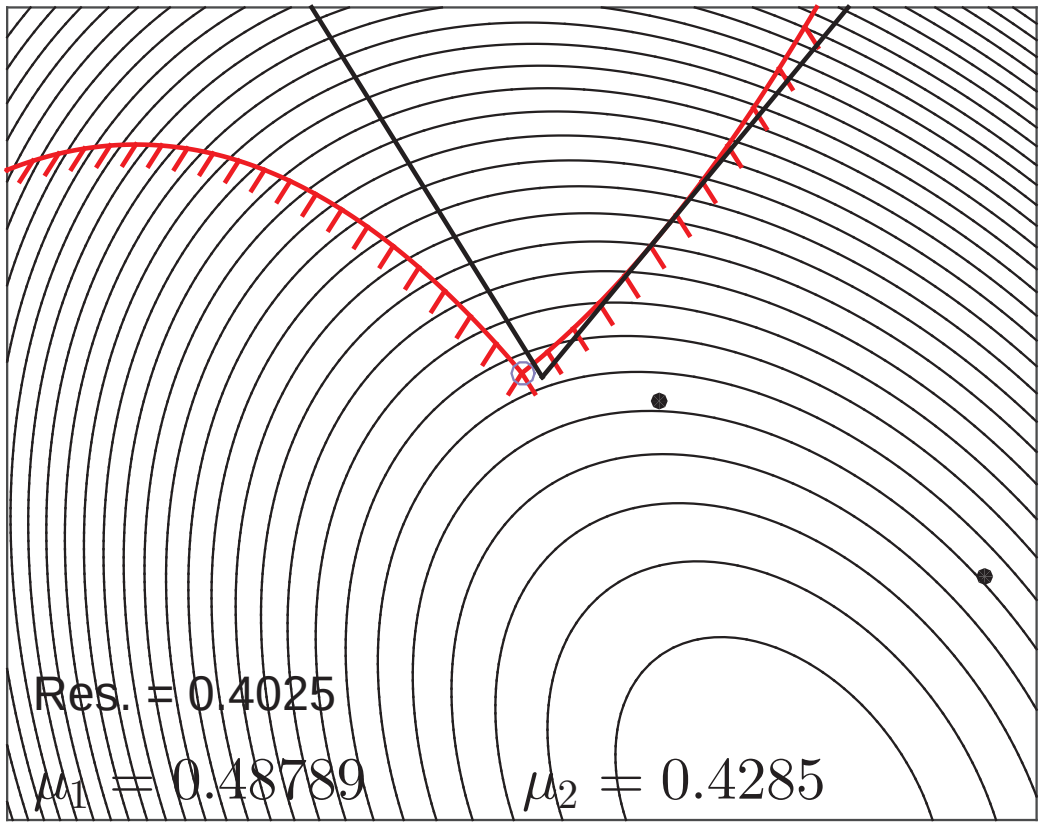
Hessian:

$$H(\mathbf{w}, \boldsymbol{\mu}) = \nabla^2 \Phi(\mathbf{w}) + \nabla^2 \left(\boldsymbol{\mu}^\top \mathbf{h}(\mathbf{w}) \right)$$

SQP - Illustration

NLP:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2} \left\| \mathbf{w} - \mathbf{w}_0 \right\|_Q^2 \\ \text{s.t.} \quad & \mathbf{h}(\mathbf{w}) \leq 0 \end{aligned}$$



QP:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2} \Delta \mathbf{w}^\top H(\mathbf{w}, \boldsymbol{\mu}) \Delta \mathbf{w} + \nabla \Phi(\mathbf{w})^\top \Delta \mathbf{w} \\ \text{s.t.} \quad & \mathbf{h}(\mathbf{w}) + \nabla \mathbf{h}(\mathbf{w})^\top \Delta \mathbf{w} \leq 0 \end{aligned}$$

Hessian:

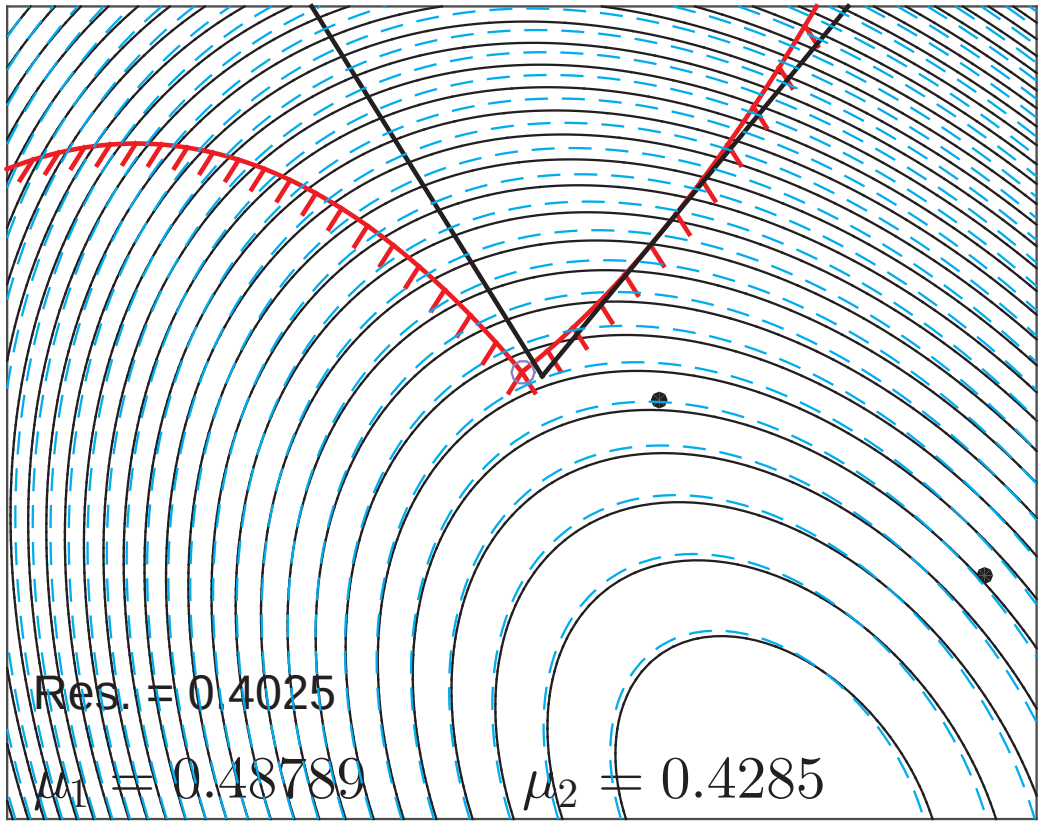
$$H(\mathbf{w}, \boldsymbol{\mu}) = \nabla^2 \Phi(\mathbf{w}) + \nabla^2 \left(\boldsymbol{\mu}^\top \mathbf{h}(\mathbf{w}) \right)$$

SQP - Illustration

NLP:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2} \left\| \mathbf{w} - \mathbf{w}_0 \right\|_Q^2 \\ \text{s.t.} \quad & \mathbf{h}(\mathbf{w}) \leq 0 \end{aligned}$$

Contours of QP cost



QP:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2} \Delta\mathbf{w}^\top H(\mathbf{w}, \boldsymbol{\mu}) \Delta\mathbf{w} + \nabla\Phi(\mathbf{w})^\top \Delta\mathbf{w} \\ \text{s.t.} \quad & \mathbf{h}(\mathbf{w}) + \nabla\mathbf{h}(\mathbf{w})^\top \Delta\mathbf{w} \leq 0 \end{aligned}$$

Hessian:

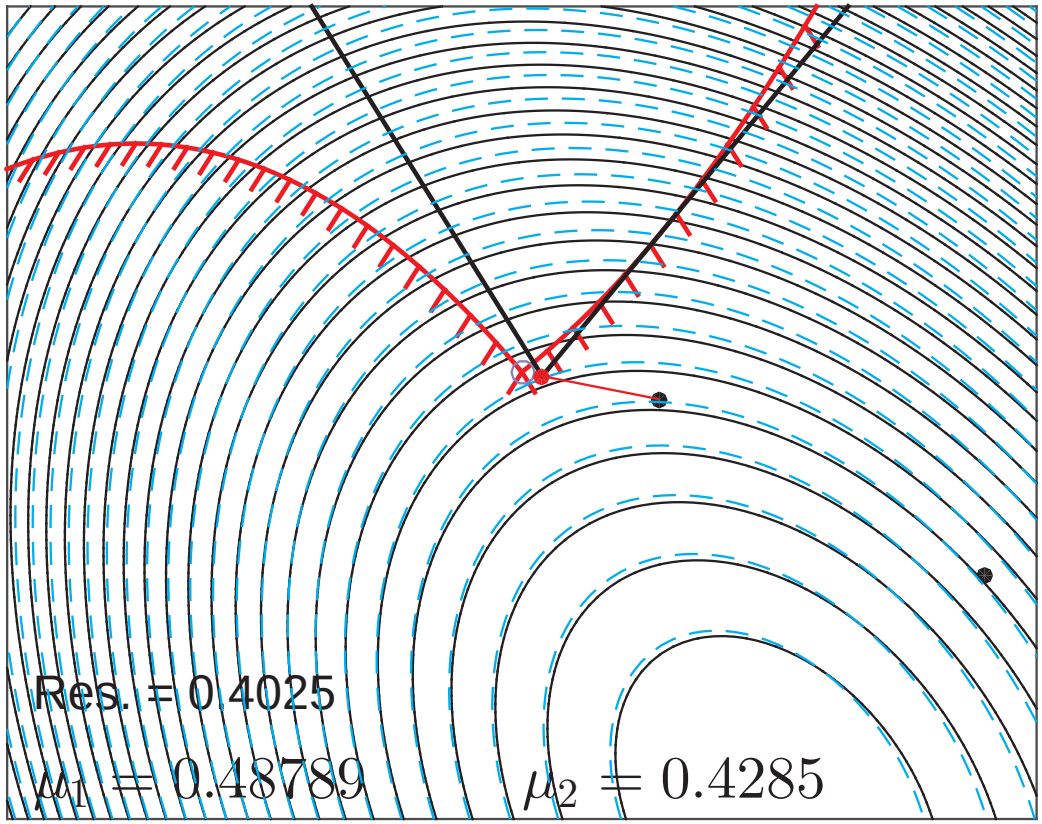
$$H(\mathbf{w}, \boldsymbol{\mu}) = \nabla^2\Phi(\mathbf{w}) + \nabla^2\left(\boldsymbol{\mu}^\top \mathbf{h}(\mathbf{w})\right)$$

SQP - Illustration

NLP:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2} \left\| \mathbf{w} - \mathbf{w}_0 \right\|_Q^2 \\ \text{s.t.} \quad & \mathbf{h}(\mathbf{w}) \leq 0 \end{aligned}$$

Step with t = 1



QP:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2} \Delta\mathbf{w}^\top H(\mathbf{w}, \boldsymbol{\mu}) \Delta\mathbf{w} + \nabla\Phi(\mathbf{w})^\top \Delta\mathbf{w} \\ \text{s.t.} \quad & \mathbf{h}(\mathbf{w}) + \nabla\mathbf{h}(\mathbf{w})^\top \Delta\mathbf{w} \leq 0 \end{aligned}$$

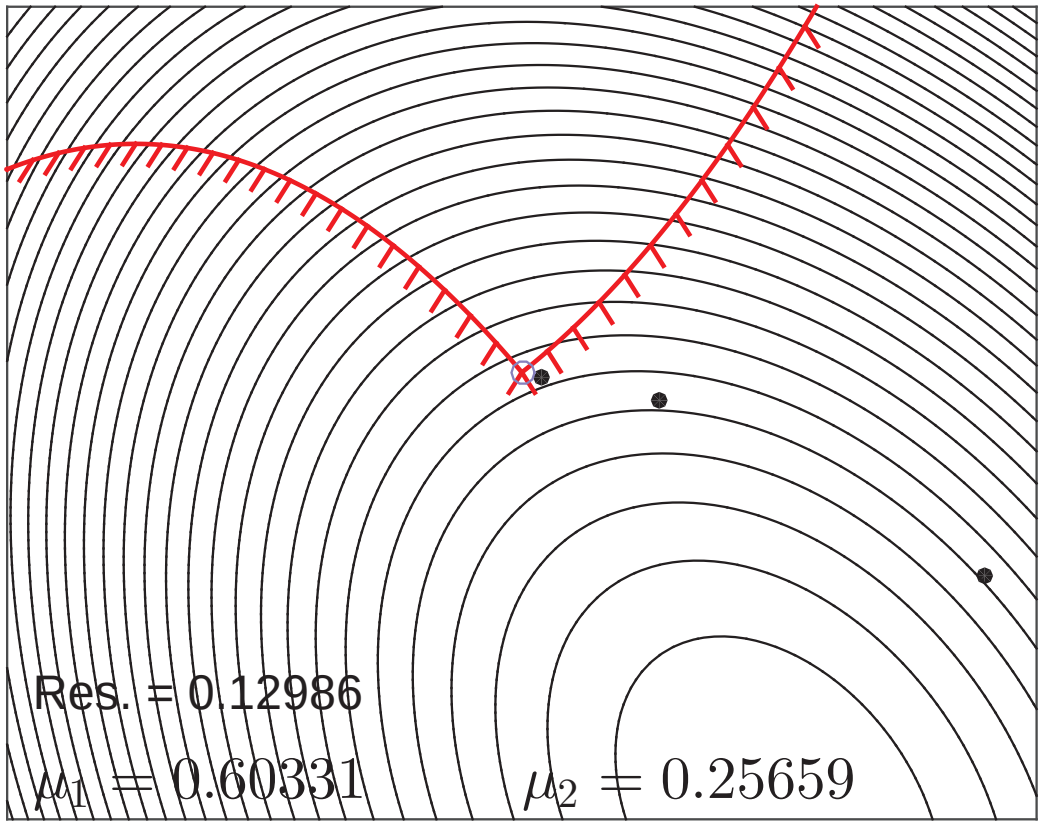
Hessian:

$$H(\mathbf{w}, \boldsymbol{\mu}) = \nabla^2\Phi(\mathbf{w}) + \nabla^2\left(\boldsymbol{\mu}^\top \mathbf{h}(\mathbf{w})\right)$$

SQP - Illustration

NLP:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2} \left\| \mathbf{w} - \mathbf{w}_0 \right\|_Q^2 \\ \text{s.t.} \quad & \mathbf{h}(\mathbf{w}) \leq 0 \end{aligned}$$



QP:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2} \Delta \mathbf{w}^\top H(\mathbf{w}, \boldsymbol{\mu}) \Delta \mathbf{w} + \nabla \Phi(\mathbf{w})^\top \Delta \mathbf{w} \\ \text{s.t.} \quad & \mathbf{h}(\mathbf{w}) + \nabla \mathbf{h}(\mathbf{w})^\top \Delta \mathbf{w} \leq 0 \end{aligned}$$

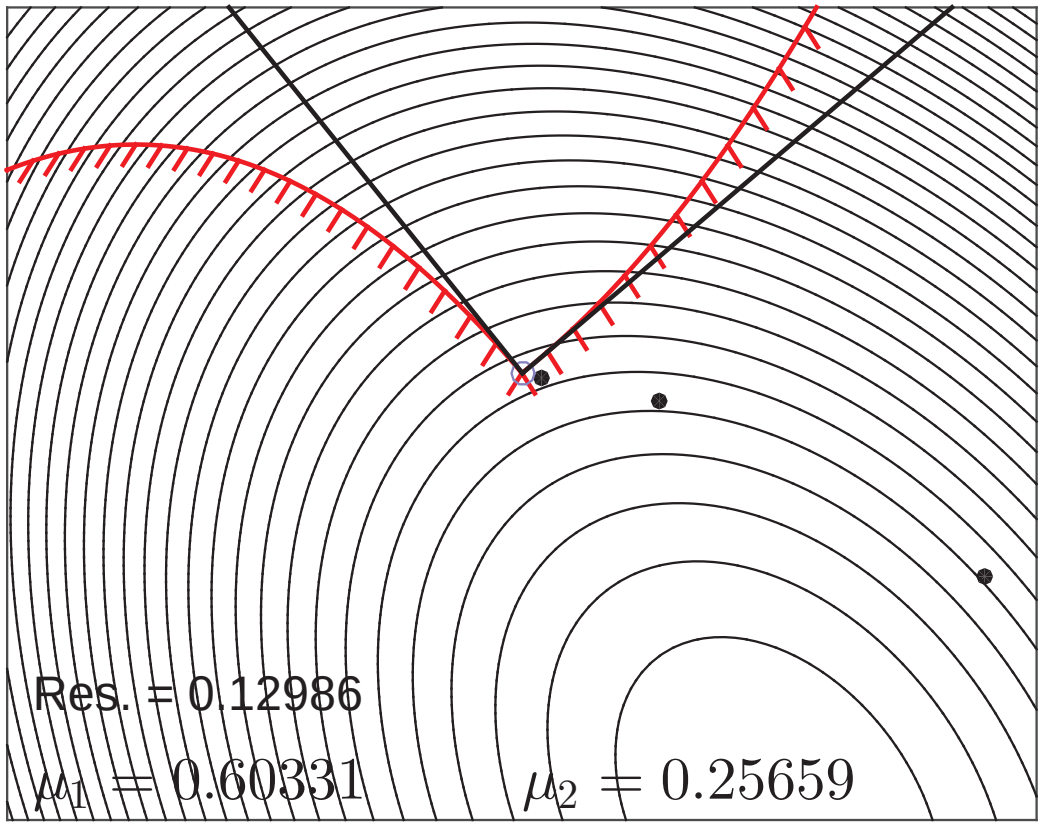
Hessian:

$$H(\mathbf{w}, \boldsymbol{\mu}) = \nabla^2 \Phi(\mathbf{w}) + \nabla^2 \left(\boldsymbol{\mu}^\top \mathbf{h}(\mathbf{w}) \right)$$

SQP - Illustration

NLP:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2} \left\| \mathbf{w} - \mathbf{w}_0 \right\|_Q^2 \\ \text{s.t.} \quad & \mathbf{h}(\mathbf{w}) \leq 0 \end{aligned}$$



QP:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2} \Delta \mathbf{w}^\top H(\mathbf{w}, \boldsymbol{\mu}) \Delta \mathbf{w} + \nabla \Phi(\mathbf{w})^\top \Delta \mathbf{w} \\ \text{s.t.} \quad & \mathbf{h}(\mathbf{w}) + \nabla \mathbf{h}(\mathbf{w})^\top \Delta \mathbf{w} \leq 0 \end{aligned}$$

Hessian:

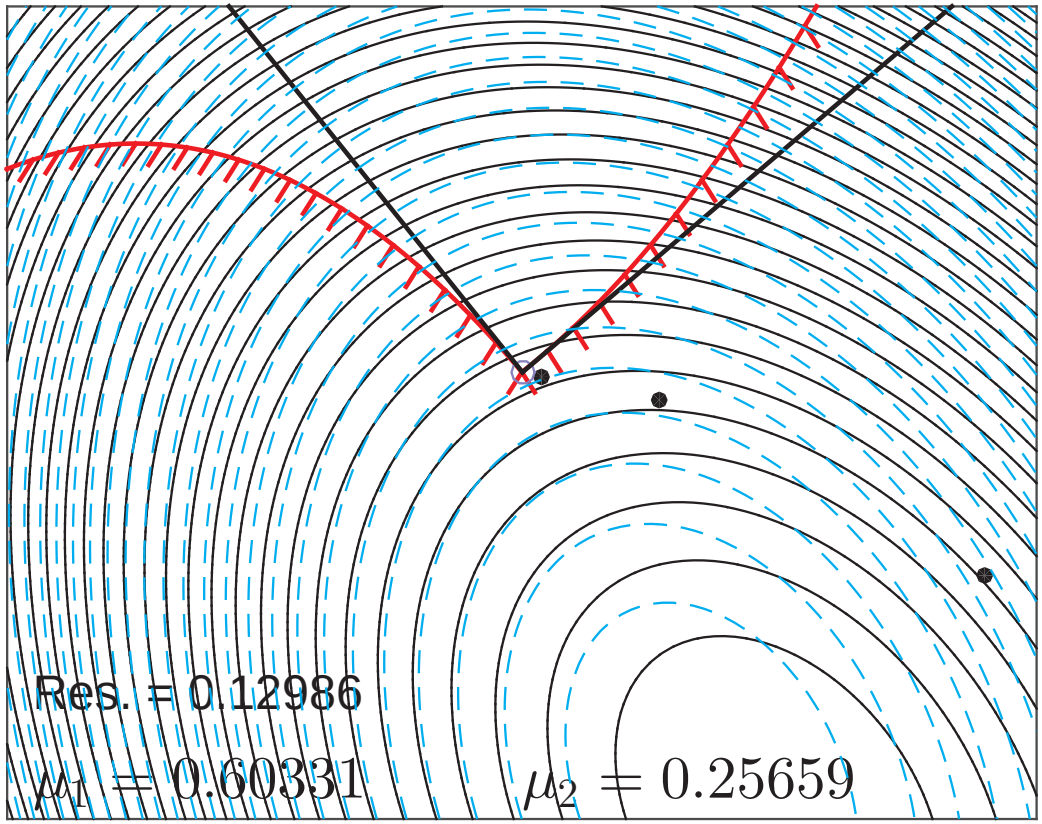
$$H(\mathbf{w}, \boldsymbol{\mu}) = \nabla^2 \Phi(\mathbf{w}) + \nabla^2 \left(\boldsymbol{\mu}^\top \mathbf{h}(\mathbf{w}) \right)$$

SQP - Illustration

NLP:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2} \left\| \mathbf{w} - \mathbf{w}_0 \right\|_Q^2 \\ \text{s.t.} \quad & \mathbf{h}(\mathbf{w}) \leq 0 \end{aligned}$$

Contours of QP cost



QP:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2} \Delta \mathbf{w}^\top H(\mathbf{w}, \boldsymbol{\mu}) \Delta \mathbf{w} + \nabla \Phi(\mathbf{w})^\top \Delta \mathbf{w} \\ \text{s.t.} \quad & \mathbf{h}(\mathbf{w}) + \nabla \mathbf{h}(\mathbf{w})^\top \Delta \mathbf{w} \leq 0 \end{aligned}$$

Hessian:

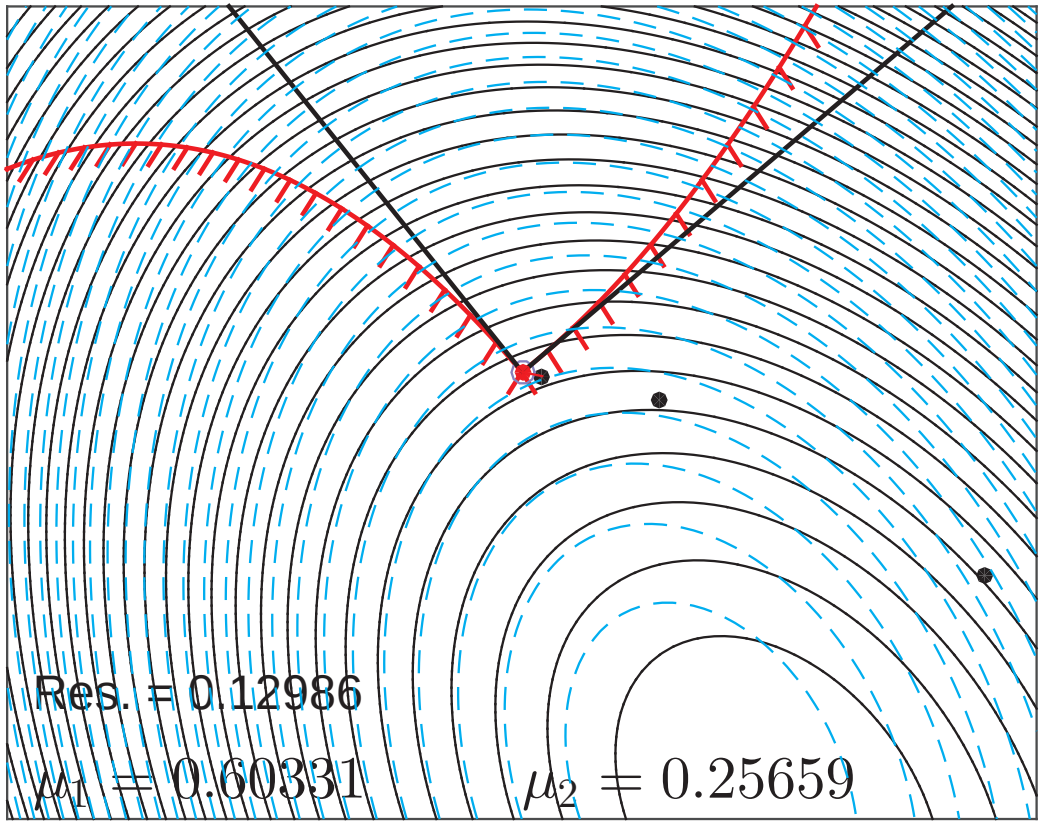
$$H(\mathbf{w}, \boldsymbol{\mu}) = \nabla^2 \Phi(\mathbf{w}) + \nabla^2 \left(\boldsymbol{\mu}^\top \mathbf{h}(\mathbf{w}) \right)$$

SQP - Illustration

NLP:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2} \left\| \mathbf{w} - \mathbf{w}_0 \right\|_Q^2 \\ \text{s.t.} \quad & \mathbf{h}(\mathbf{w}) \leq 0 \end{aligned}$$

Step with t = 1



QP:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2} \Delta \mathbf{w}^\top H(\mathbf{w}, \boldsymbol{\mu}) \Delta \mathbf{w} + \nabla \Phi(\mathbf{w})^\top \Delta \mathbf{w} \\ \text{s.t.} \quad & \mathbf{h}(\mathbf{w}) + \nabla \mathbf{h}(\mathbf{w})^\top \Delta \mathbf{w} \leq 0 \end{aligned}$$

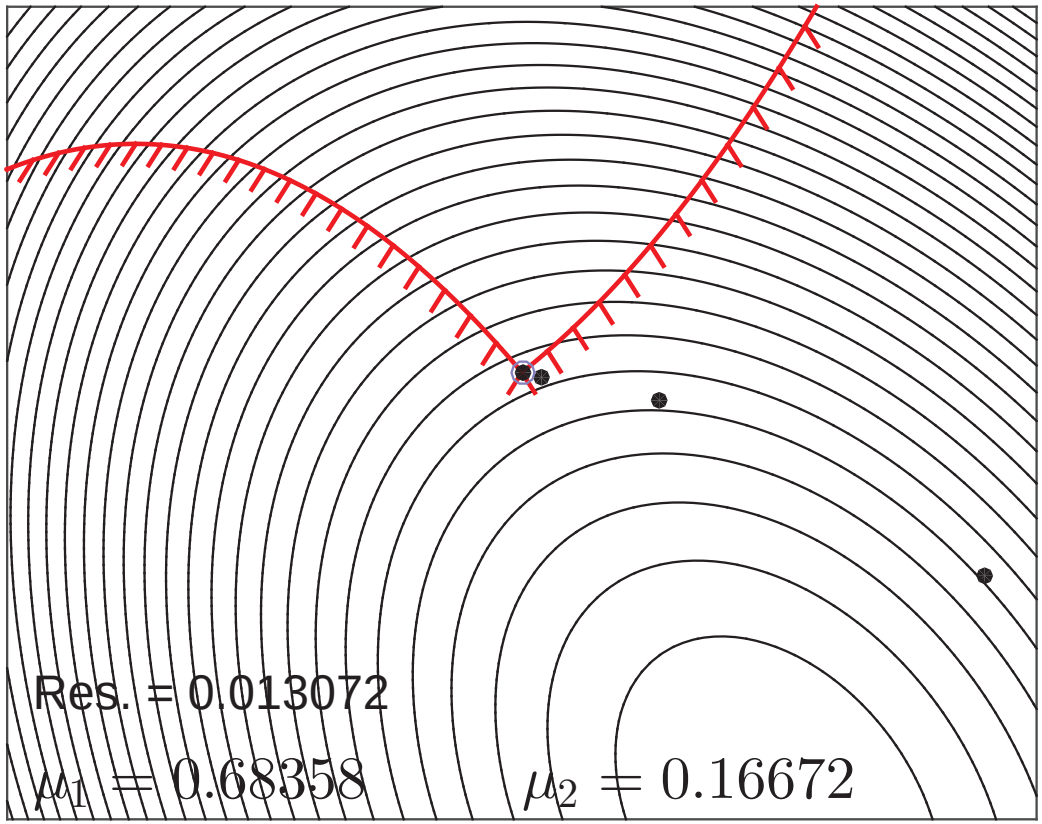
Hessian:

$$H(\mathbf{w}, \boldsymbol{\mu}) = \nabla^2 \Phi(\mathbf{w}) + \nabla^2 \left(\boldsymbol{\mu}^\top \mathbf{h}(\mathbf{w}) \right)$$

SQP - Illustration

NLP:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2} \left\| \mathbf{w} - \mathbf{w}_0 \right\|_Q^2 \\ \text{s.t.} \quad & \mathbf{h}(\mathbf{w}) \leq 0 \end{aligned}$$



QP:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2} \Delta \mathbf{w}^\top H(\mathbf{w}, \boldsymbol{\mu}) \Delta \mathbf{w} + \nabla \Phi(\mathbf{w})^\top \Delta \mathbf{w} \\ \text{s.t.} \quad & \mathbf{h}(\mathbf{w}) + \nabla \mathbf{h}(\mathbf{w})^\top \Delta \mathbf{w} \leq 0 \end{aligned}$$

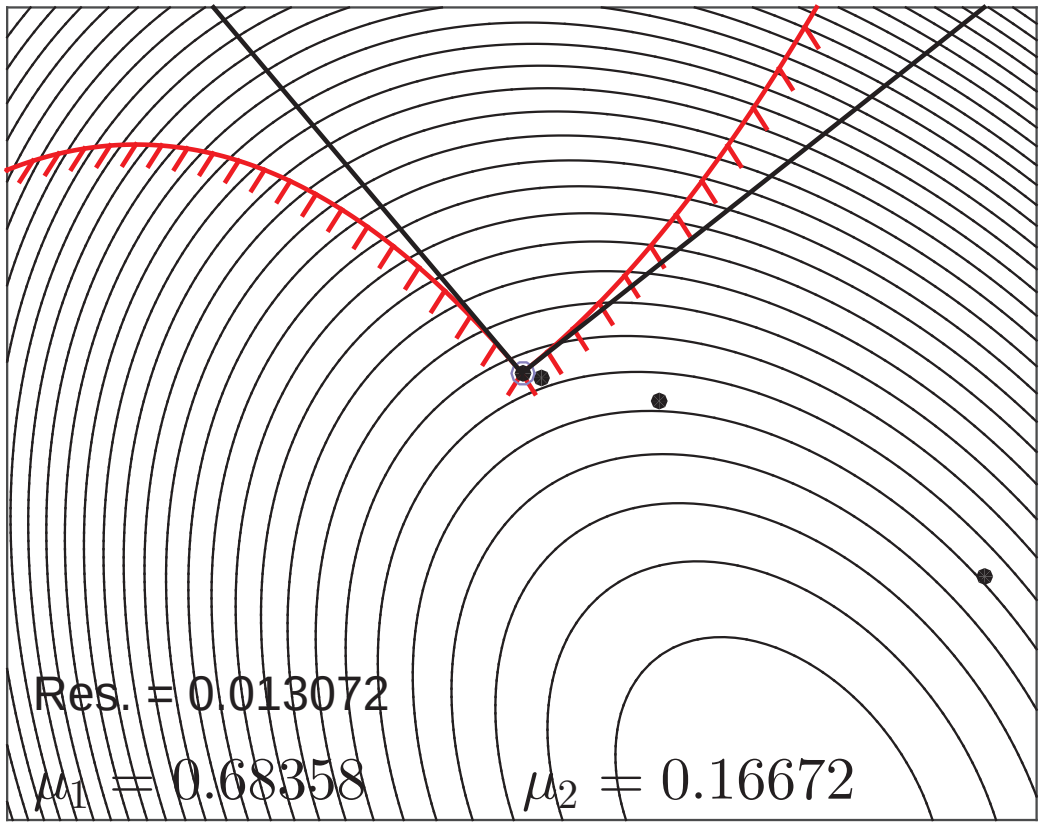
Hessian:

$$H(\mathbf{w}, \boldsymbol{\mu}) = \nabla^2 \Phi(\mathbf{w}) + \nabla^2 \left(\boldsymbol{\mu}^\top \mathbf{h}(\mathbf{w}) \right)$$

SQP - Illustration

NLP:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2} \left\| \mathbf{w} - \mathbf{w}_0 \right\|_Q^2 \\ \text{s.t.} \quad & \mathbf{h}(\mathbf{w}) \leq 0 \end{aligned}$$



QP:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2} \Delta \mathbf{w}^\top H(\mathbf{w}, \boldsymbol{\mu}) \Delta \mathbf{w} + \nabla \Phi(\mathbf{w})^\top \Delta \mathbf{w} \\ \text{s.t.} \quad & \mathbf{h}(\mathbf{w}) + \nabla \mathbf{h}(\mathbf{w})^\top \Delta \mathbf{w} \leq 0 \end{aligned}$$

Hessian:

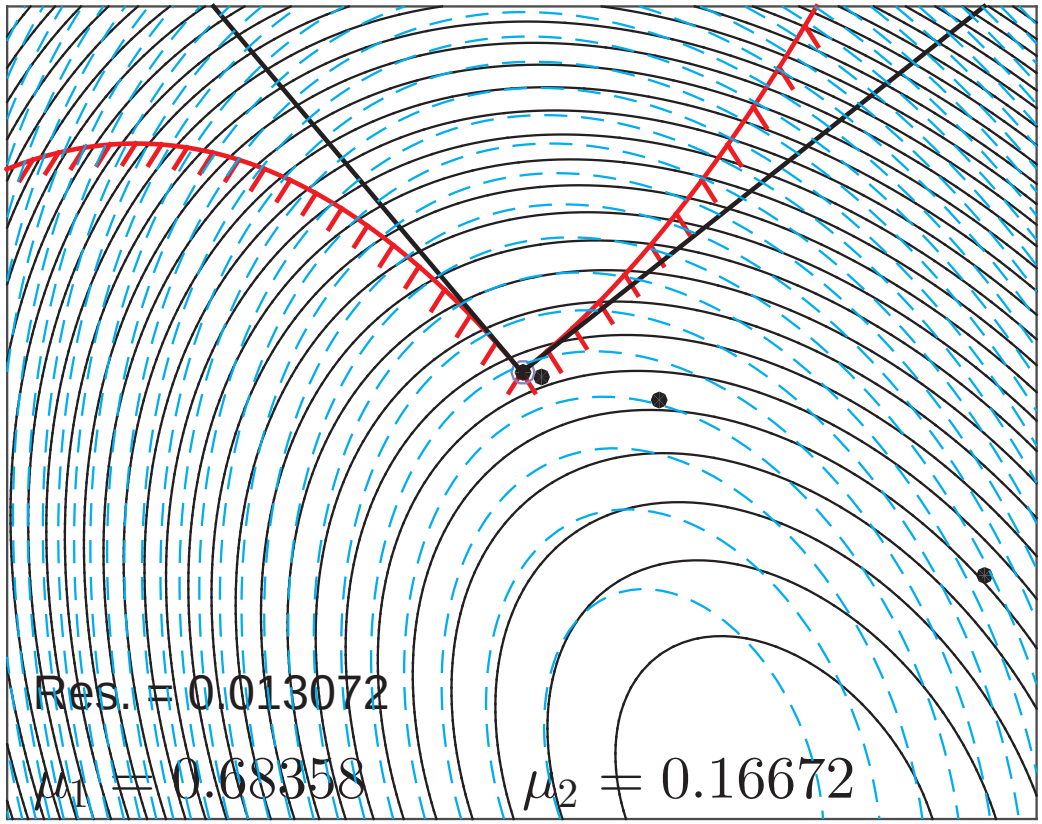
$$H(\mathbf{w}, \boldsymbol{\mu}) = \nabla^2 \Phi(\mathbf{w}) + \nabla^2 \left(\boldsymbol{\mu}^\top \mathbf{h}(\mathbf{w}) \right)$$

SQP - Illustration

NLP:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2} \left\| \mathbf{w} - \mathbf{w}_0 \right\|_Q^2 \\ \text{s.t.} \quad & \mathbf{h}(\mathbf{w}) \leq 0 \end{aligned}$$

Contours of QP cost



QP:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2} \Delta \mathbf{w}^\top H(\mathbf{w}, \boldsymbol{\mu}) \Delta \mathbf{w} + \nabla \Phi(\mathbf{w})^\top \Delta \mathbf{w} \\ \text{s.t.} \quad & \mathbf{h}(\mathbf{w}) + \nabla \mathbf{h}(\mathbf{w})^\top \Delta \mathbf{w} \leq 0 \end{aligned}$$

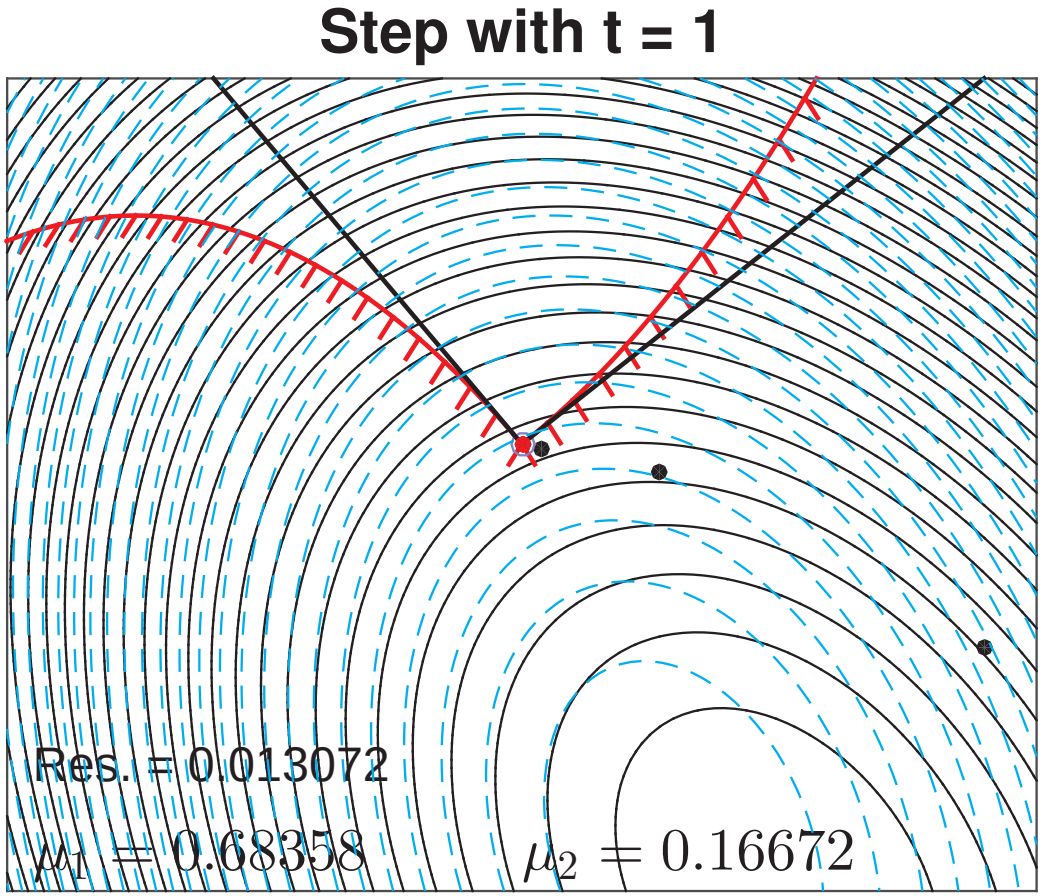
Hessian:

$$H(\mathbf{w}, \boldsymbol{\mu}) = \nabla^2 \Phi(\mathbf{w}) + \nabla^2 \left(\boldsymbol{\mu}^\top \mathbf{h}(\mathbf{w}) \right)$$

SQP - Illustration

NLP:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2} \left\| \mathbf{w} - \mathbf{w}_0 \right\|_Q^2 \\ \text{s.t.} \quad & \mathbf{h}(\mathbf{w}) \leq 0 \end{aligned}$$



QP:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2} \Delta \mathbf{w}^\top H(\mathbf{w}, \boldsymbol{\mu}) \Delta \mathbf{w} + \nabla \Phi(\mathbf{w})^\top \Delta \mathbf{w} \\ \text{s.t.} \quad & \mathbf{h}(\mathbf{w}) + \nabla \mathbf{h}(\mathbf{w})^\top \Delta \mathbf{w} \leq 0 \end{aligned}$$

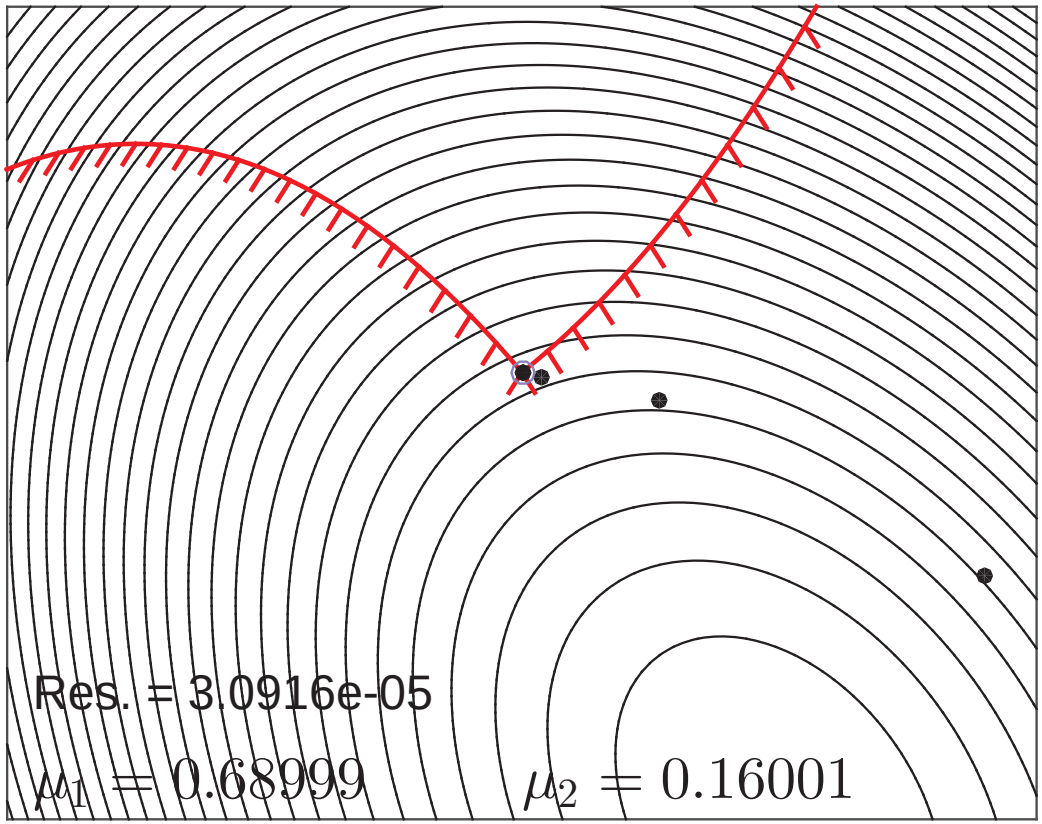
Hessian:

$$H(\mathbf{w}, \boldsymbol{\mu}) = \nabla^2 \Phi(\mathbf{w}) + \nabla^2 \left(\boldsymbol{\mu}^\top \mathbf{h}(\mathbf{w}) \right)$$

SQP - Illustration

NLP:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2} \left\| \mathbf{w} - \mathbf{w}_0 \right\|_Q^2 \\ \text{s.t.} \quad & \mathbf{h}(\mathbf{w}) \leq 0 \end{aligned}$$



QP:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2} \Delta \mathbf{w}^\top H(\mathbf{w}, \boldsymbol{\mu}) \Delta \mathbf{w} + \nabla \Phi(\mathbf{w})^\top \Delta \mathbf{w} \\ \text{s.t.} \quad & \mathbf{h}(\mathbf{w}) + \nabla \mathbf{h}(\mathbf{w})^\top \Delta \mathbf{w} \leq 0 \end{aligned}$$

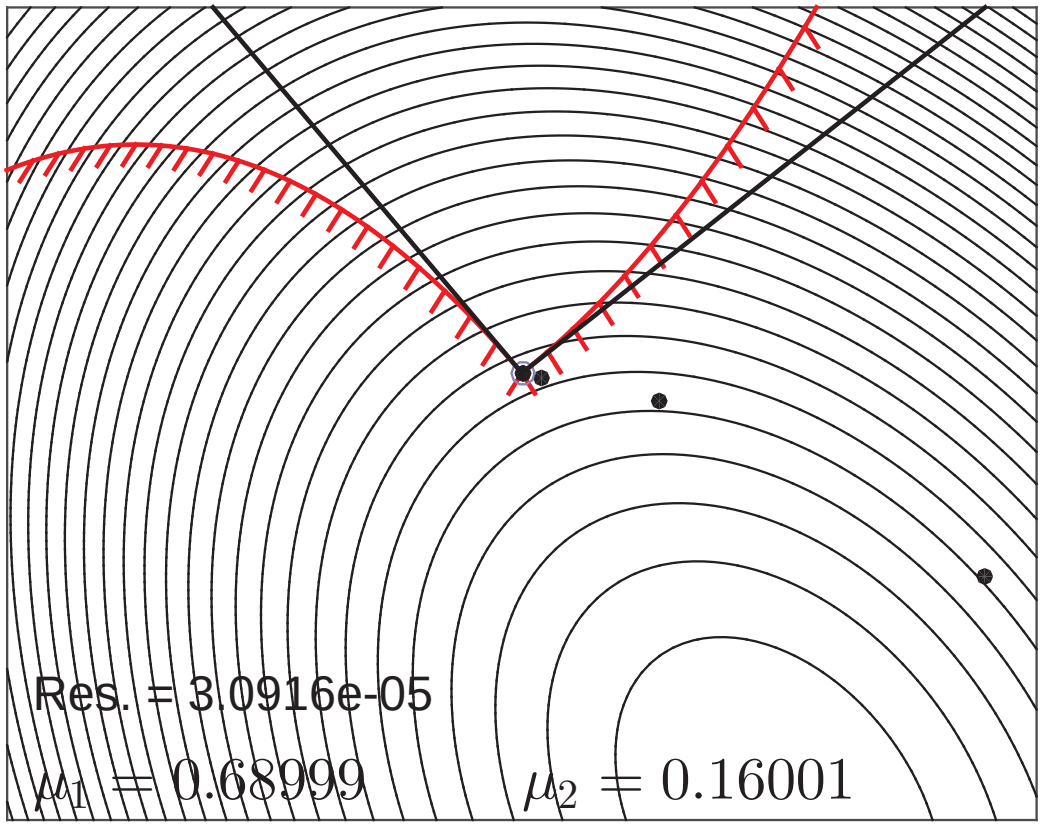
Hessian:

$$H(\mathbf{w}, \boldsymbol{\mu}) = \nabla^2 \Phi(\mathbf{w}) + \nabla^2 \left(\boldsymbol{\mu}^\top \mathbf{h}(\mathbf{w}) \right)$$

SQP - Illustration

NLP:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2} \left\| \mathbf{w} - \mathbf{w}_0 \right\|_Q^2 \\ \text{s.t.} \quad & \mathbf{h}(\mathbf{w}) \leq 0 \end{aligned}$$



QP:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2} \Delta \mathbf{w}^\top H(\mathbf{w}, \boldsymbol{\mu}) \Delta \mathbf{w} + \nabla \Phi(\mathbf{w})^\top \Delta \mathbf{w} \\ \text{s.t.} \quad & \mathbf{h}(\mathbf{w}) + \nabla \mathbf{h}(\mathbf{w})^\top \Delta \mathbf{w} \leq 0 \end{aligned}$$

Hessian:

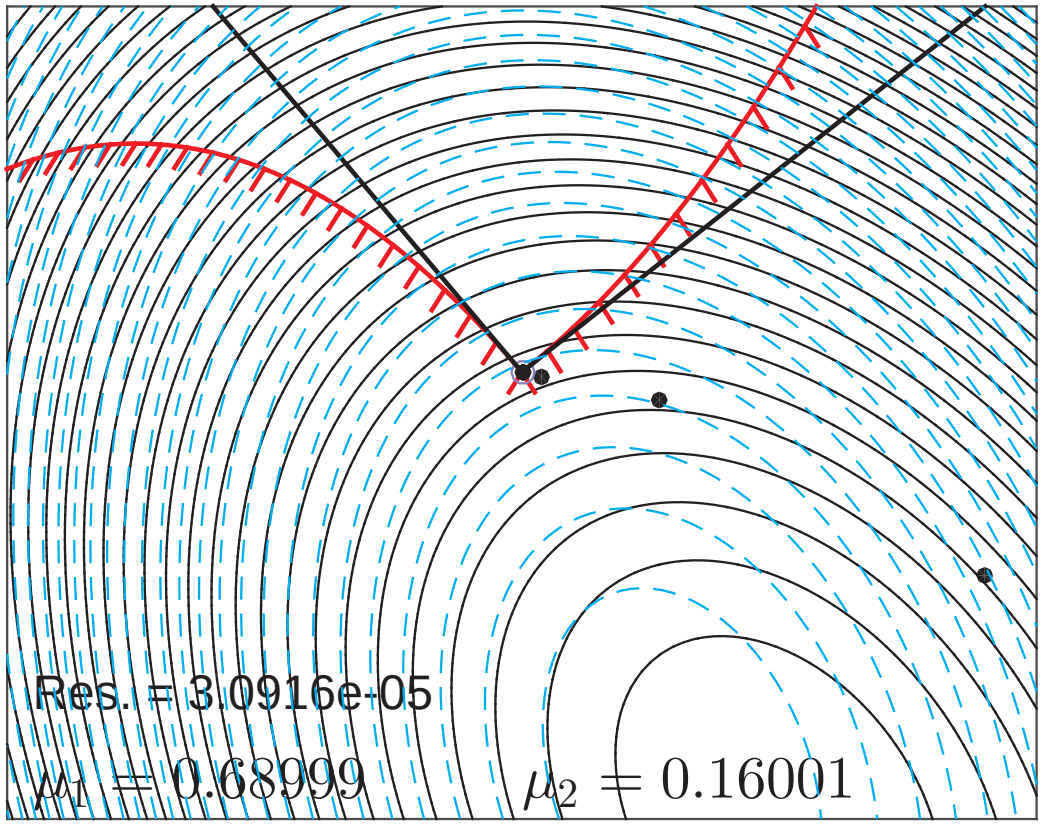
$$H(\mathbf{w}, \boldsymbol{\mu}) = \nabla^2 \Phi(\mathbf{w}) + \nabla^2 \left(\boldsymbol{\mu}^\top \mathbf{h}(\mathbf{w}) \right)$$

SQP - Illustration

NLP:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2} \left\| \mathbf{w} - \mathbf{w}_0 \right\|_Q^2 \\ \text{s.t.} \quad & \mathbf{h}(\mathbf{w}) \leq 0 \end{aligned}$$

Contours of QP cost



QP:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2} \Delta \mathbf{w}^\top H(\mathbf{w}, \boldsymbol{\mu}) \Delta \mathbf{w} + \nabla \Phi(\mathbf{w})^\top \Delta \mathbf{w} \\ \text{s.t.} \quad & \mathbf{h}(\mathbf{w}) + \nabla \mathbf{h}(\mathbf{w})^\top \Delta \mathbf{w} \leq 0 \end{aligned}$$

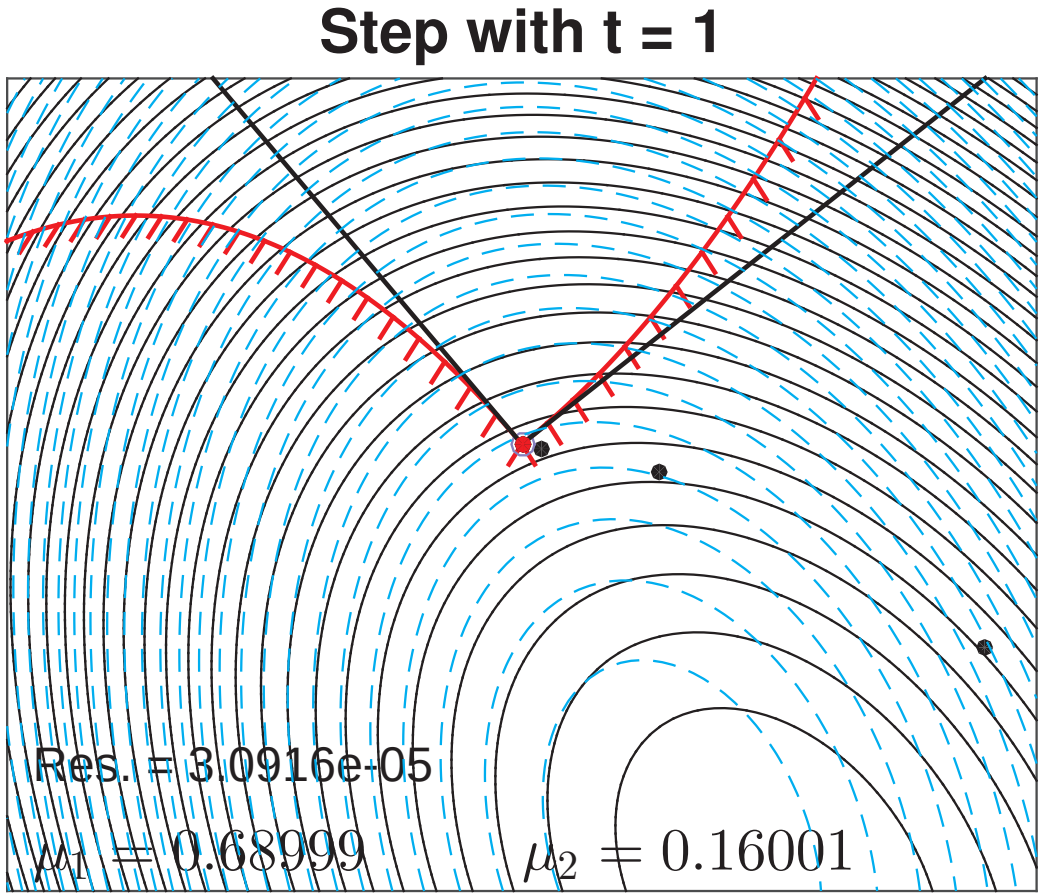
Hessian:

$$H(\mathbf{w}, \boldsymbol{\mu}) = \nabla^2 \Phi(\mathbf{w}) + \nabla^2 \left(\boldsymbol{\mu}^\top \mathbf{h}(\mathbf{w}) \right)$$

SQP - Illustration

NLP:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2} \left\| \mathbf{w} - \mathbf{w}_0 \right\|_Q^2 \\ \text{s.t.} \quad & \mathbf{h}(\mathbf{w}) \leq 0 \end{aligned}$$



QP:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2} \Delta \mathbf{w}^\top H(\mathbf{w}, \boldsymbol{\mu}) \Delta \mathbf{w} + \nabla \Phi(\mathbf{w})^\top \Delta \mathbf{w} \\ \text{s.t.} \quad & \mathbf{h}(\mathbf{w}) + \nabla \mathbf{h}(\mathbf{w})^\top \Delta \mathbf{w} \leq 0 \end{aligned}$$

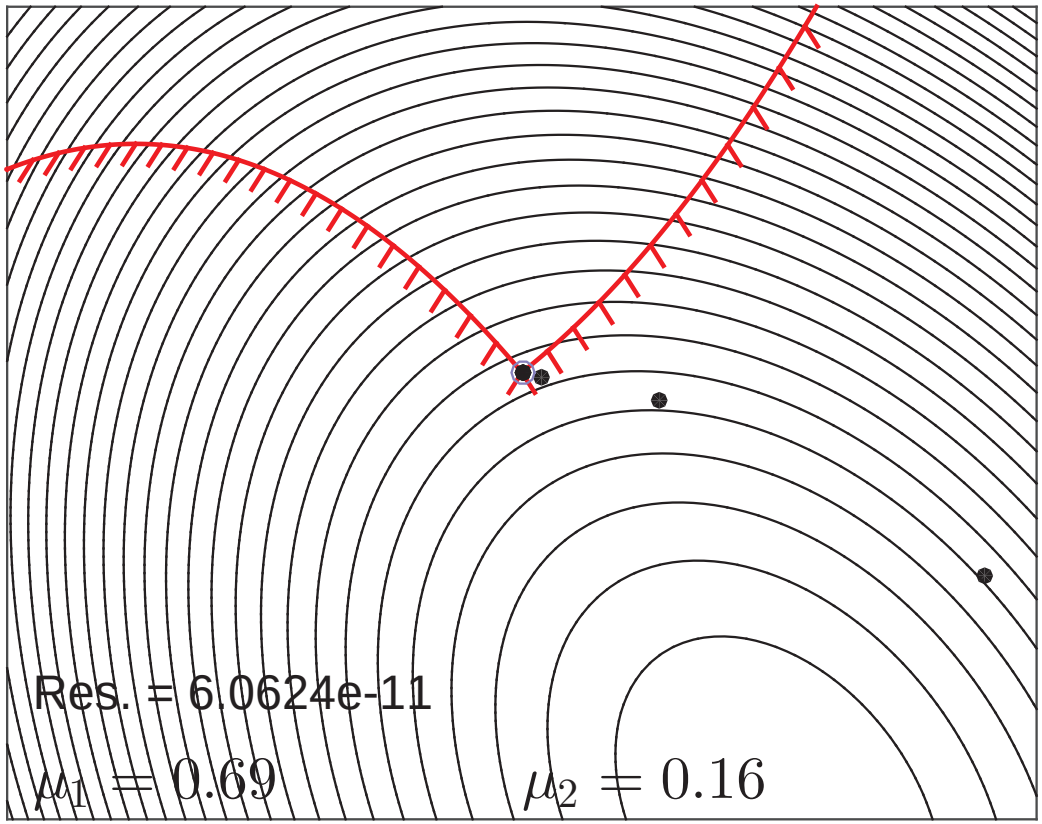
Hessian:

$$H(\mathbf{w}, \boldsymbol{\mu}) = \nabla^2 \Phi(\mathbf{w}) + \nabla^2 \left(\boldsymbol{\mu}^\top \mathbf{h}(\mathbf{w}) \right)$$

SQP - Illustration

NLP:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2} \left\| \mathbf{w} - \mathbf{w}_0 \right\|_Q^2 \\ \text{s.t.} \quad & \mathbf{h}(\mathbf{w}) \leq 0 \end{aligned}$$



QP:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2} \Delta \mathbf{w}^\top H(\mathbf{w}, \boldsymbol{\mu}) \Delta \mathbf{w} + \nabla \Phi(\mathbf{w})^\top \Delta \mathbf{w} \\ \text{s.t.} \quad & \mathbf{h}(\mathbf{w}) + \nabla \mathbf{h}(\mathbf{w})^\top \Delta \mathbf{w} \leq 0 \end{aligned}$$

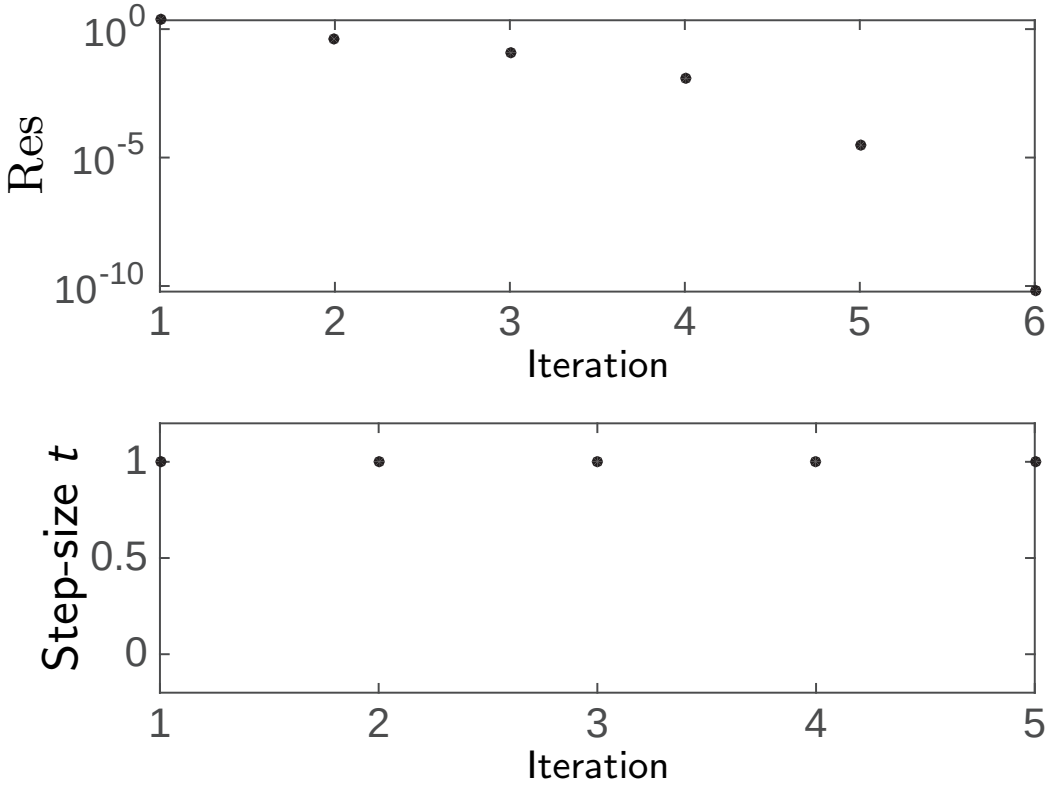
Hessian:

$$H(\mathbf{w}, \boldsymbol{\mu}) = \nabla^2 \Phi(\mathbf{w}) + \nabla^2 \left(\boldsymbol{\mu}^\top \mathbf{h}(\mathbf{w}) \right)$$

SQP - Illustration

NLP:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2} \left\| \mathbf{w} - \mathbf{w}_0 \right\|_Q^2 \\ \text{s.t.} \quad & \mathbf{h}(\mathbf{w}) \leq 0 \end{aligned}$$



QP:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2} \Delta \mathbf{w}^\top H(\mathbf{w}, \boldsymbol{\mu}) \Delta \mathbf{w} + \nabla \Phi(\mathbf{w})^\top \Delta \mathbf{w} \\ \text{s.t.} \quad & \mathbf{h}(\mathbf{w}) + \nabla \mathbf{h}(\mathbf{w})^\top \Delta \mathbf{w} \leq 0 \end{aligned}$$

Hessian:

$$H(\mathbf{w}, \boldsymbol{\mu}) = \nabla^2 \Phi(\mathbf{w}) + \nabla^2 \left(\boldsymbol{\mu}^\top \mathbf{h}(\mathbf{w}) \right)$$

Maratos effect

Consider the NLP :

$$\min_{u,v} \quad \Phi = 3v^2 - 2u$$

$$\text{s.t.} \quad g = u - v^2 = 0$$

Optimum: $\mathbf{w}^* = \begin{bmatrix} 0 & 0 \end{bmatrix}$.

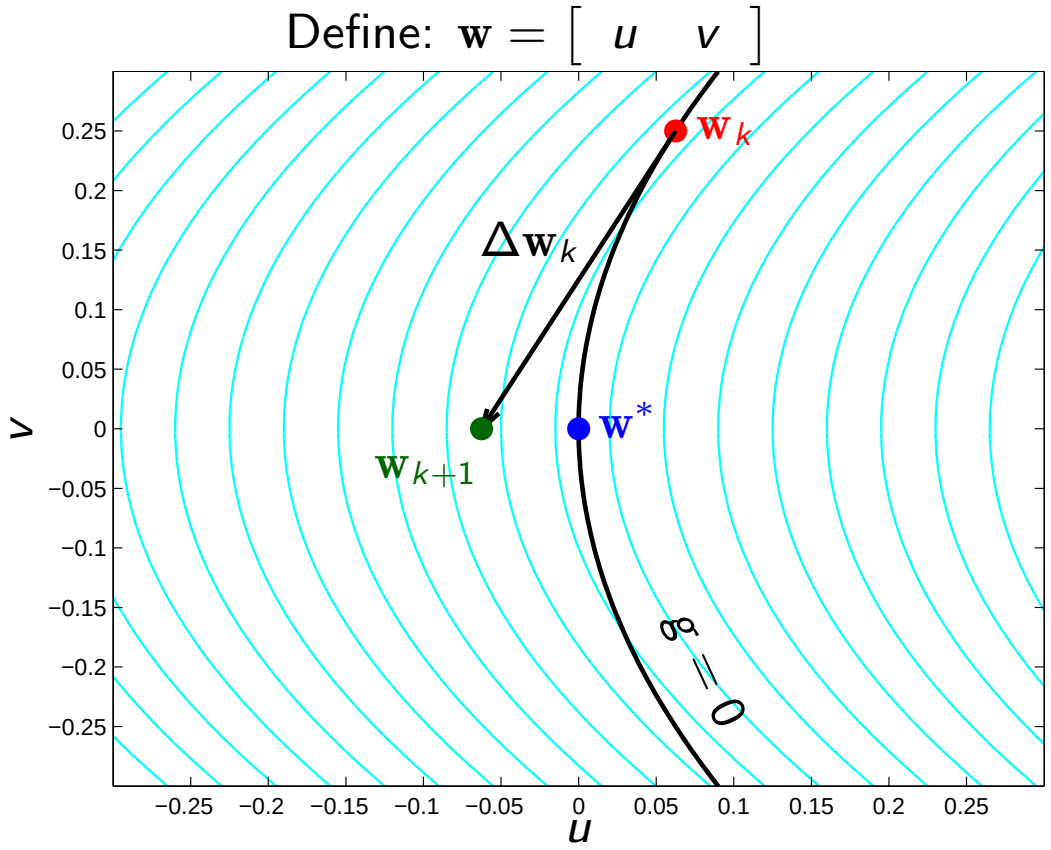
Consider the iterate:

$$\mathbf{w}_k = \begin{bmatrix} a^2 & a \end{bmatrix}$$

The Newton step is:

$$\Delta \mathbf{w}_k = -\begin{bmatrix} 2a^2 & a \end{bmatrix}$$

for $\lambda = 2$...



Maratos effect

Consider the NLP :

$$\min_{u,v} \quad \Phi = 3v^2 - 2u$$

$$\text{s.t.} \quad g = u - v^2 = 0$$

Optimum: $\mathbf{w}^* = \begin{bmatrix} 0 & 0 \end{bmatrix}$.

Consider the iterate:

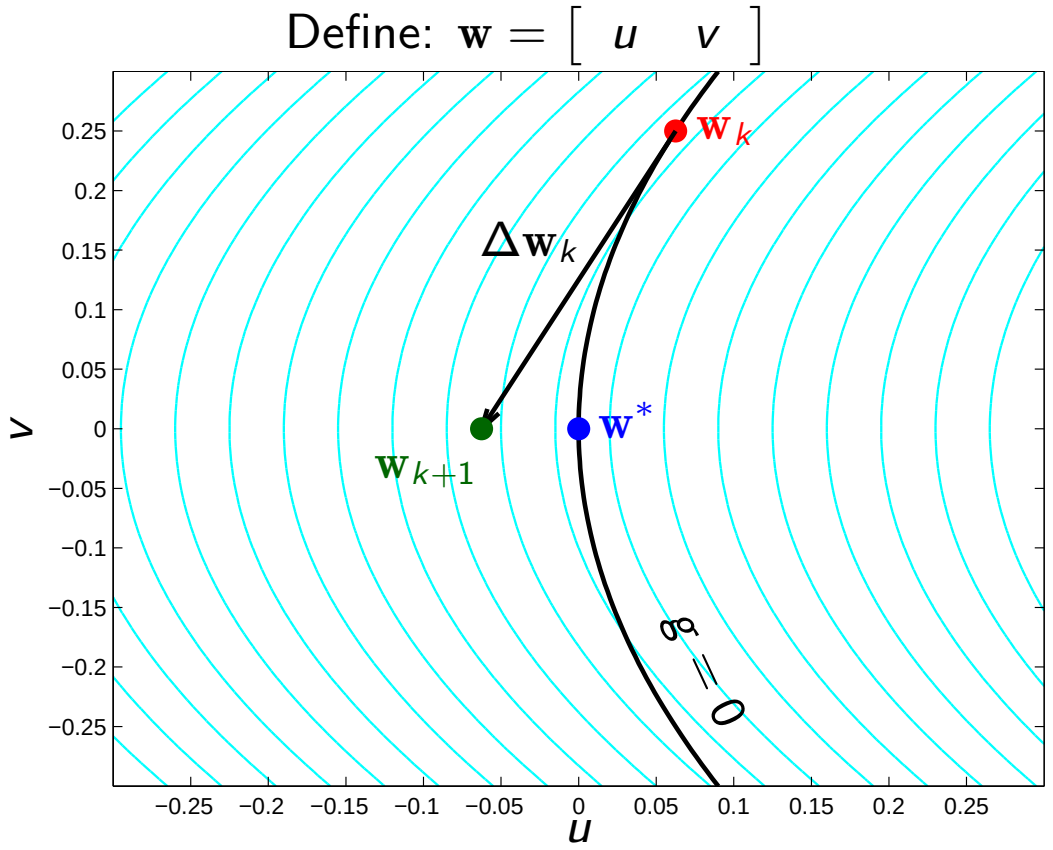
$$\mathbf{w}_k = \begin{bmatrix} a^2 & a \end{bmatrix}$$

The Newton step is:

$$\Delta \mathbf{w}_k = -\begin{bmatrix} 2a^2 & a \end{bmatrix}$$

for $\lambda = 2$...

The full Newton step:
$\mathbf{w}_{k+1} = \mathbf{w}_k + \Delta \mathbf{w}_k$
is much closer to $\mathbf{w}^*$ than $\mathbf{w}_k$.



Maratos effect

Consider the NLP :

$$\min_{u,v} \quad \Phi = 3v^2 - 2u$$

$$\text{s.t.} \quad g = u - v^2 = 0$$

Optimum: $\mathbf{w}^* = \begin{bmatrix} 0 & 0 \end{bmatrix}$.

Consider the iterate:

$$\mathbf{w}_k = \begin{bmatrix} a^2 & a \end{bmatrix}$$

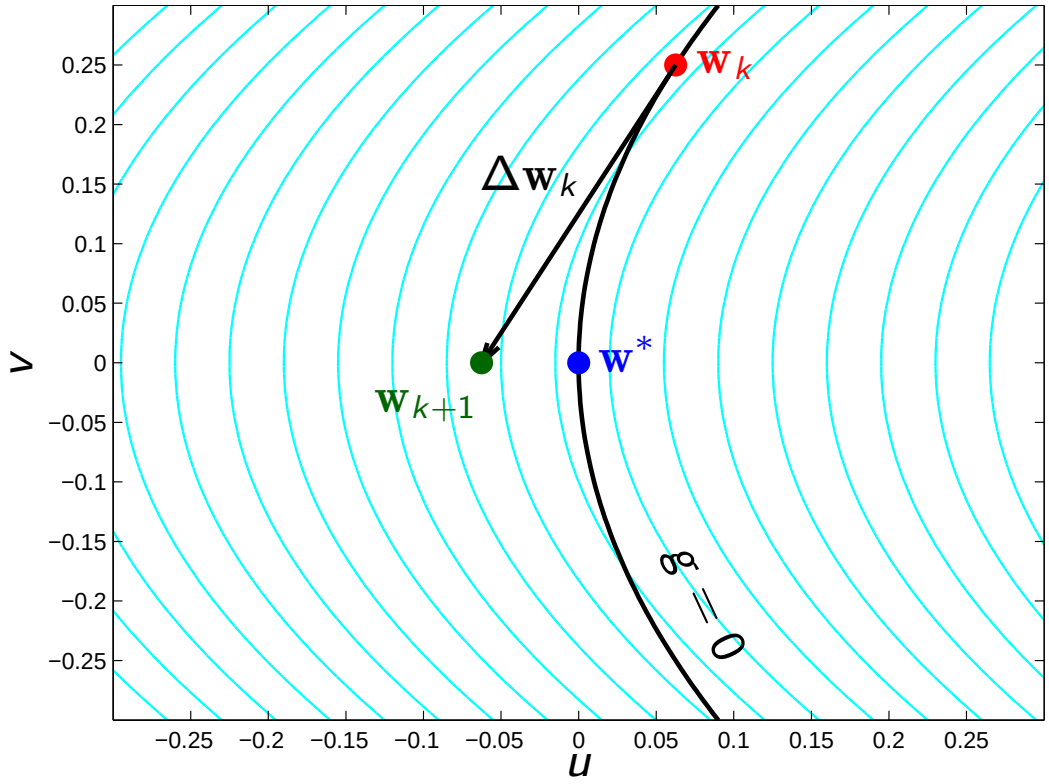
The Newton step is:

$$\Delta \mathbf{w}_k = -\begin{bmatrix} 2a^2 & a \end{bmatrix}$$

for $\lambda = 2...$

The full Newton step:
$\mathbf{w}_{k+1} = \mathbf{w}_k + \Delta \mathbf{w}_k$
is much closer to $\mathbf{w}^*$ than $\mathbf{w}_k$.

Define: $\mathbf{w} = \begin{bmatrix} u & v \end{bmatrix}$



But:
$\Phi(\mathbf{w}_{k+1}) > \Phi(\mathbf{w}_k)$
$|g(\mathbf{w}_{k+1})| > |g(\mathbf{w}_k)|$

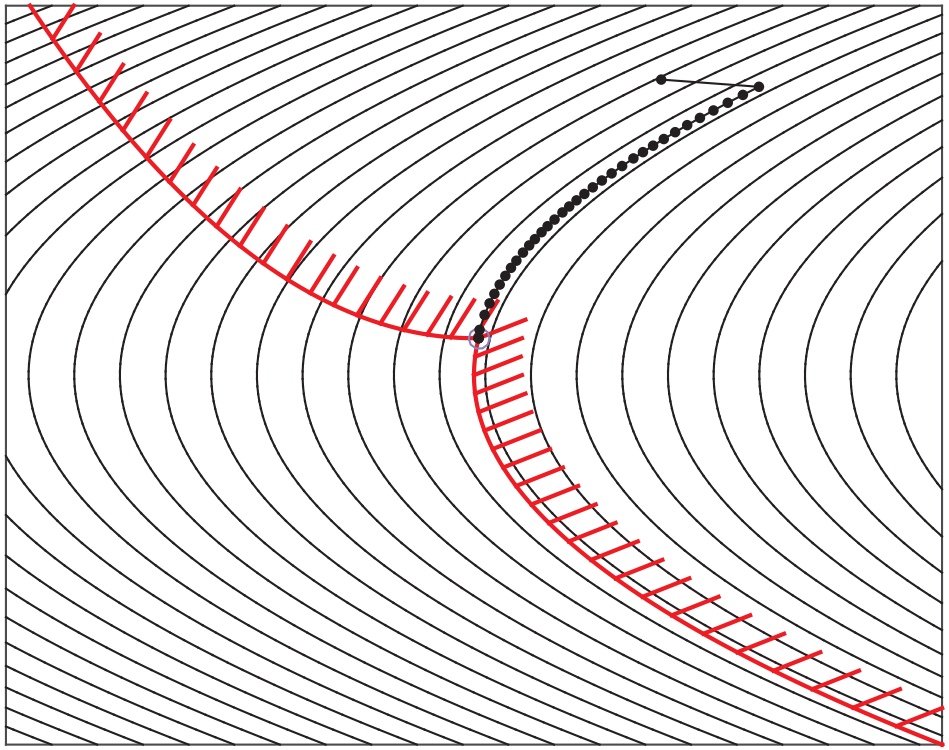
No penalty function can accept $\Delta \mathbf{w}_k$!!

Maratos effect - Illustration

Some NLPs can yield "creeping" convergence

NLP:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \Phi(\mathbf{w}) \\ \text{s.t.} \quad & \mathbf{h}(\mathbf{w}) \leq 0 \end{aligned}$$

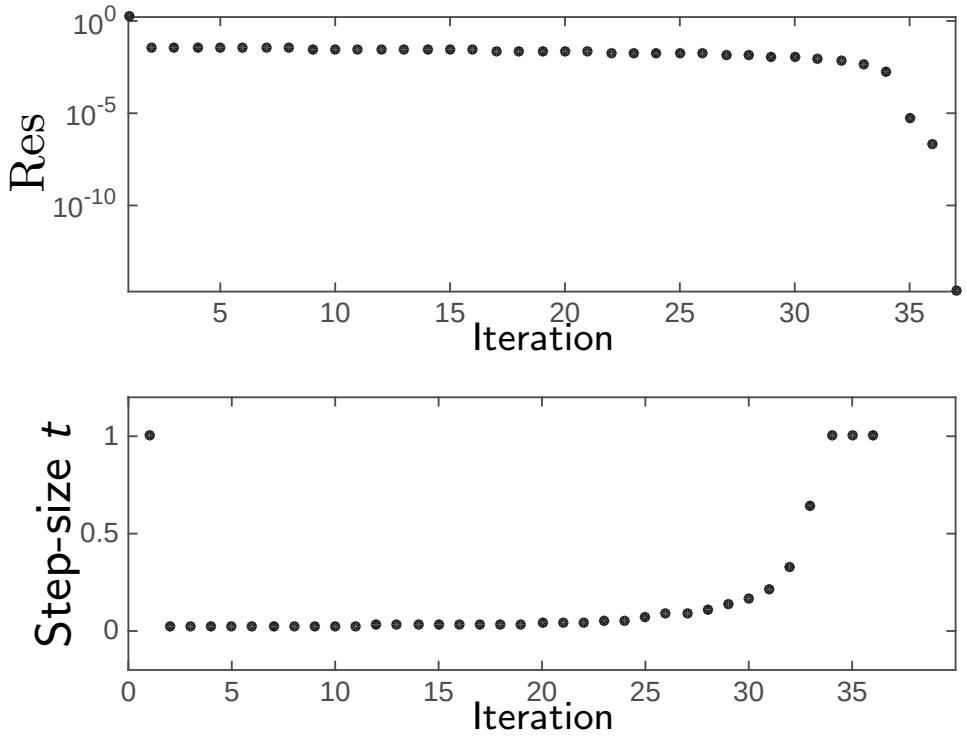


Maratos effect - Illustration

Some NLPs can yield "creeping" convergence

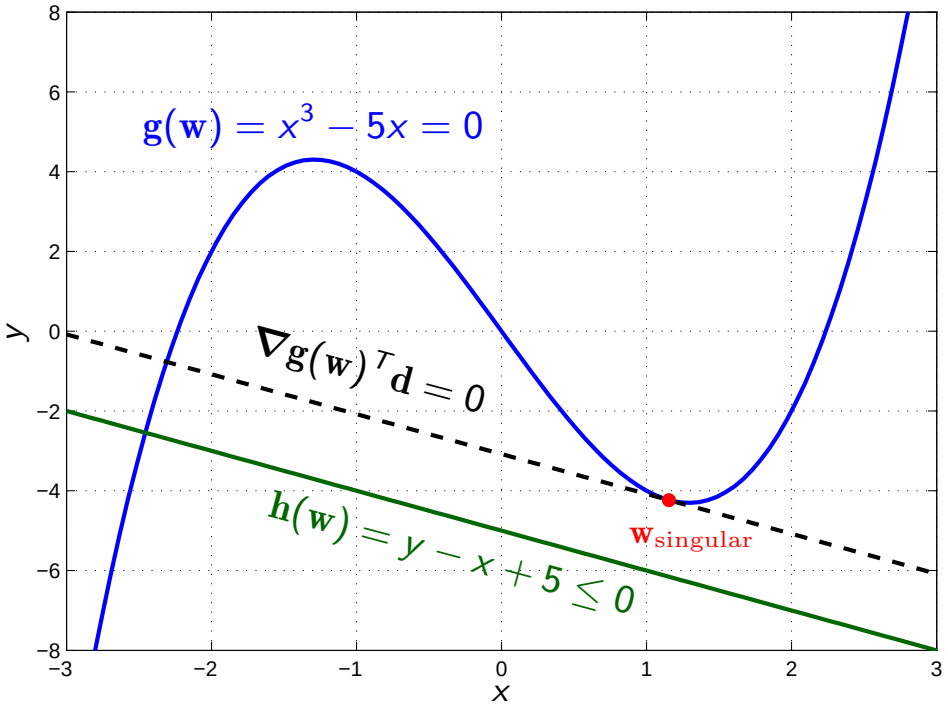
NLP:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \Phi(\mathbf{w}) \\ \text{s.t.} \quad & \mathbf{h}(\mathbf{w}) \leq 0 \end{aligned}$$



where $\mathrm{Res} = \left\| \begin{array}{c} \nabla \mathcal{L} \\ \max(\mathbf{h}, 0) \end{array} \right\|_{\infty}$

Failure of the Newton methods - Convergence to infeasible points



Problem:

$$\begin{array}{ll} \min\limits_{\mathbf{w}} & x^2 + y^2 \\ \text{s.t.} & x^3 - 5x = 0 \\ & y - x + 5 \leq 0 \end{array}$$

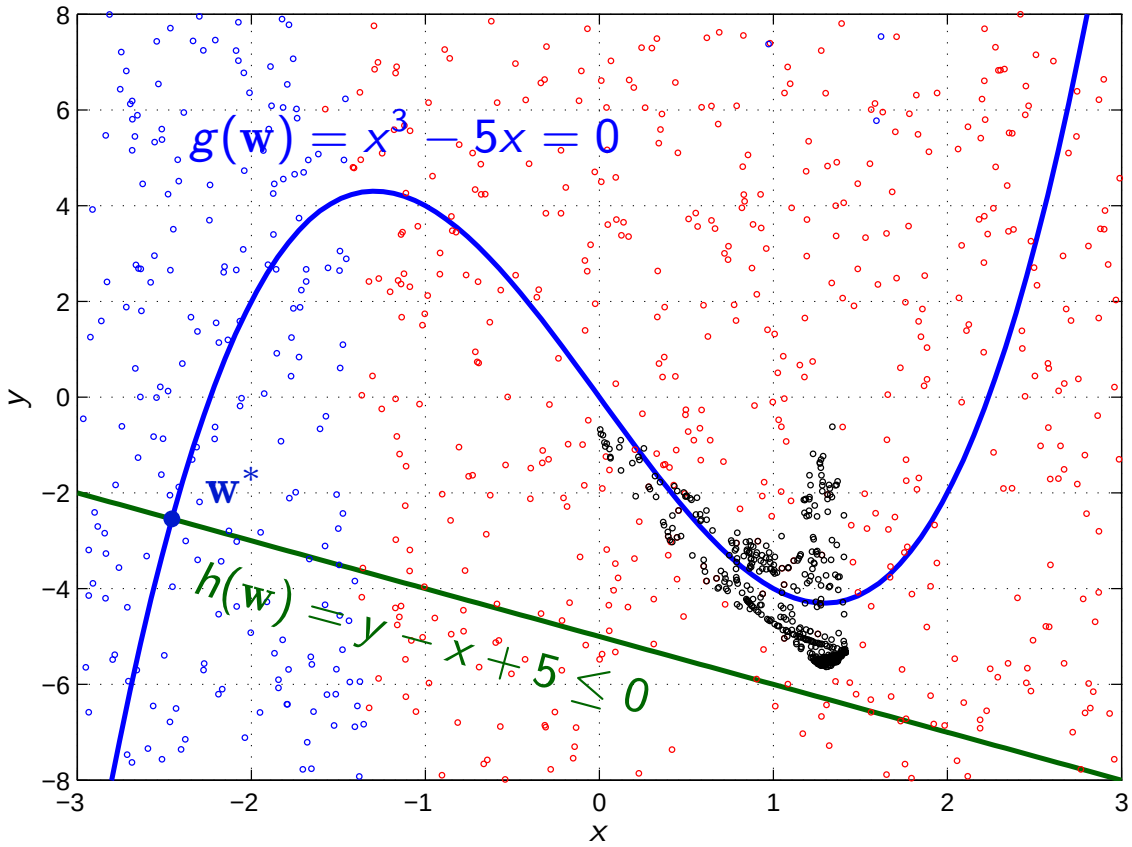
Vectors $\nabla \mathbf{g}(\mathbf{w})$ and $\nabla \mathbf{h}(\mathbf{w})$ are not Linearly Independent for some $\mathbf{w}$, i.e.

$\nexists \Delta \mathbf{w}$ such that:

$$\begin{bmatrix} \nabla \mathbf{g}(\mathbf{w})^\top \\ \nabla \mathbf{h}(\mathbf{w})^\top \end{bmatrix} \Delta \mathbf{w} + \begin{bmatrix} \mathbf{g}(\mathbf{w}) \\ \mathbf{h}(\mathbf{w}) \end{bmatrix} = 0$$

There is no feasible direction $\Delta \mathbf{w}$

Failure of the Newton methods - Convergence to infeasible points



Problem:

$$\begin{aligned} \min_{x,y} \quad & x^2 + y^2 \\ \text{s.t.} \quad & x^3 - 5x = 0 \\ & y - x + 5 \leq 0 \end{aligned}$$

Red dots: failed starting points $\rightarrow$ black dots

Blue dots: succesfull starting points $\rightarrow (-2.46, -2.54)$