In case you missed it - Who am I ?

Name:	Sébastien Gros
Nationality:	Swiss
Residence:	Göteborg, Sweden
Affiliation:	Chalmers University of Technology
Department:	Signals & Systems
Position:	Assistant Professor
Email:	grosse@chalmers.se
Tel.	+46 31 772 15 55

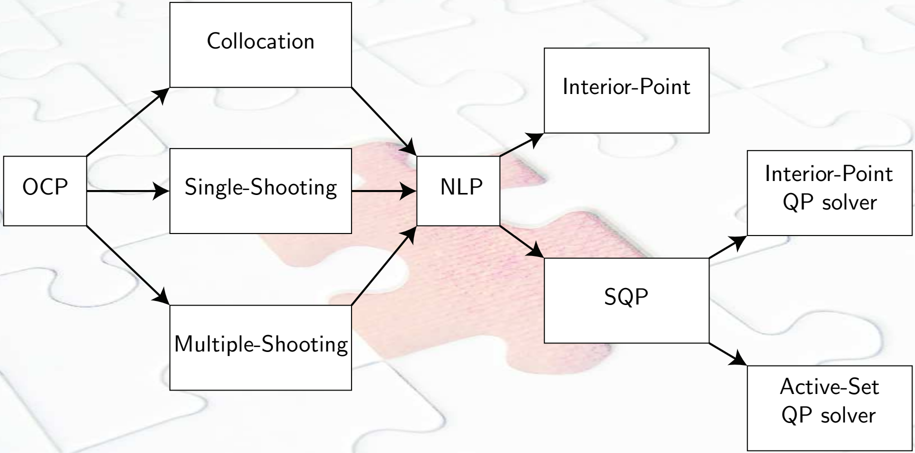
Recent research topics: distributed & parallelized methods for optimal control, estimation & system identification, NMPC & Economic NMPC, optimal control for complex mechanical systems, integrators for real-time optimal control, robust optimal control, aerospace applications, airborne wind energy, wind turbine control, smart grids, traffic control

Numerical Optimal Control with DAEs
Lecture 5: Newton method & SQP

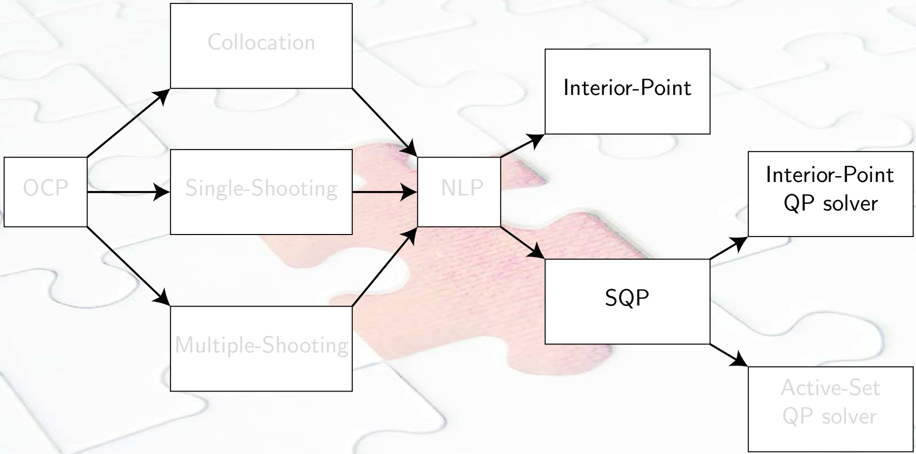
Sébastien Gros

AWESCO PhD course

Survival map of Direct Optimal Control

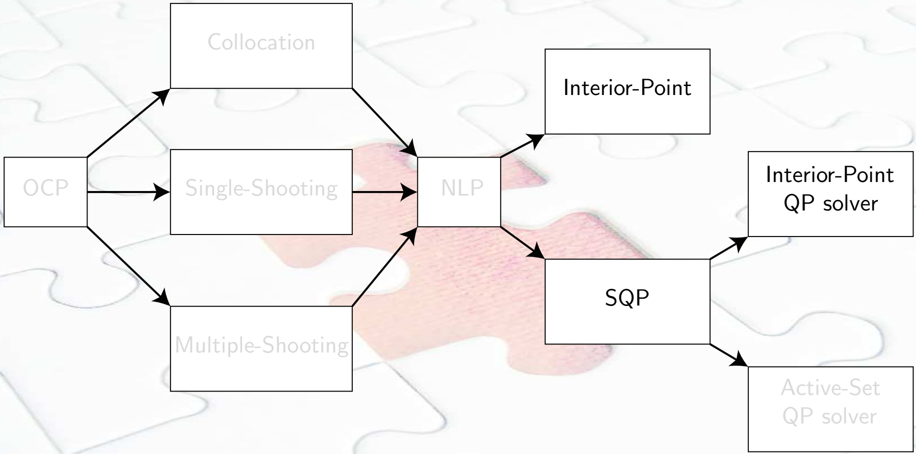


Survival map of Direct Optimal Control



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

Outline

KKT point

Consider the NLP problem:

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

KKT point

Consider the NLP problem:

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

A point $\{\mathbf{w}^*, \boldsymbol{\mu}^*, \boldsymbol{\lambda}^*\}$ is called a KKT point if it satisfies:

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

KKT point

Consider the NLP problem:

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

A point $\{\mathbf{w}^*, \boldsymbol{\mu}^*, \boldsymbol{\lambda}^*\}$ is called a KKT point if it satisfies:

$$\boxed{\textbf{Dual Feasibility:} \qquad \nabla_{\mathbf{w}}\mathcal{L}(\mathbf{w}^*, \boldsymbol{\mu}^*, \boldsymbol{\lambda}^*) = 0, \quad \boldsymbol{\mu}^* \geq 0,}$$

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

KKT point

Consider the NLP problem:

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

A point $\{\mathbf{w}^*, \boldsymbol{\mu}^*, \boldsymbol{\lambda}^*\}$ is called a KKT point if it satisfies:

$$\boxed{\begin{array}{rl} \textbf{Dual Feasibility:} & \nabla_{\mathbf{w}}\mathcal{L}(\mathbf{w}^*, \boldsymbol{\mu}^*, \boldsymbol{\lambda}^*) = 0, \quad \boldsymbol{\mu}^* \geq 0, \\ \textbf{Primal Feasibility:} & \mathbf{g}(\mathbf{w}^*) = 0, \quad \mathbf{h}(\mathbf{w}^*) \leq 0, \end{array}}$$

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

KKT point

Consider the NLP problem:

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

A point $\{\mathbf{w}^*, \boldsymbol{\mu}^*, \boldsymbol{\lambda}^*\}$ is called a KKT point if it satisfies:

$$\boxed{\begin{array}{rl} \textbf{Dual Feasibility:} & \nabla_{\mathbf{w}}\mathcal{L}\left(\mathbf{w}^*, \boldsymbol{\mu}^*, \boldsymbol{\lambda}^*\right) = 0, \quad \boldsymbol{\mu}^* \geq 0, \\ \textbf{Primal Feasibility:} & \mathbf{g}\left(\mathbf{w}^*\right) = 0, \quad \mathbf{h}\left(\mathbf{w}^*\right) \leq 0, \\ \textbf{Complementary Slackness:} & \boldsymbol{\mu}_i^* \mathbf{h}_i(\mathbf{w}^*) = 0, \quad \forall\, i \end{array}}$$

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

KKT point

Consider the NLP problem:

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

A point $\{\mathbf{w}^*, \boldsymbol{\mu}^*, \boldsymbol{\lambda}^*\}$ is called a KKT point if it satisfies:

$$\boxed{\begin{array}{rl} \textbf{Dual Feasibility:} & \nabla_{\mathbf{w}}\mathcal{L}(\mathbf{w}^*, \boldsymbol{\mu}^*, \boldsymbol{\lambda}^*) = 0, \quad \boldsymbol{\mu}^* \geq 0, \\ \textbf{Primal Feasibility:} & \mathbf{g}(\mathbf{w}^*) = 0, \quad \mathbf{h}(\mathbf{w}^*) \leq 0, \\ \textbf{Complementary Slackness:} & \boldsymbol{\mu}_i^* \mathbf{h}_i(\mathbf{w}^*) = 0, \quad \forall\, i \end{array}}$$

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

Optimality conditions for NLP with equality and/or inequality constraints:

KKT point

Consider the NLP problem:

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

A point $\{\mathbf{w}^*, \boldsymbol{\mu}^*, \boldsymbol{\lambda}^*\}$ is called a KKT point if it satisfies:

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

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

Optimality conditions for NLP with equality and/or inequality constraints:

- **1st-Order Necessary Conditions**: A (local) optimum $\mathbf{w}^\star$ satisfying LICQ of a (differentiable) NLP corresponds to a unique KKT point

KKT point

Consider the NLP problem:

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

A point $\{\mathbf{w}^*, \boldsymbol{\mu}^*, \boldsymbol{\lambda}^*\}$ is called a KKT point if it satisfies:

$$\boxed{\begin{array}{rl} \textbf{Dual Feasibility:} & \nabla_{\mathbf{w}}\mathcal{L}(\mathbf{w}^*, \boldsymbol{\mu}^*, \boldsymbol{\lambda}^*) = 0, \quad \boldsymbol{\mu}^* \geq 0, \\ \textbf{Primal Feasibility:} & \mathbf{g}(\mathbf{w}^*) = 0, \quad \mathbf{h}(\mathbf{w}^*) \leq 0, \\ \textbf{Complementary Slackness:} & \boldsymbol{\mu}_i^* \mathbf{h}_i(\mathbf{w}^*) = 0, \quad \forall\, i \end{array}}$$

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

Optimality conditions for NLP with equality and/or inequality constraints:

- **1st-Order Necessary Conditions**: A (local) optimum $\mathbf{w}^\star$ satisfying LICQ of a (differentiable) NLP corresponds to a unique KKT point
- **2nd-Order Sufficient Conditions** require positivity of the Hessian $\nabla^2_{\mathbf{w}}\mathcal{L}$ in all critical feasible directions at the solution

KKT point

Consider the NLP problem:

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

Most NLP solvers are in essence "KKT solvers"

A point $\{\mathbf{w}^*, \boldsymbol{\mu}^*, \boldsymbol{\lambda}^*\}$ is called a KKT point if it satisfies:

$$\boxed{\begin{array}{rl}\textbf{Dual Feasibility:} & \nabla_{\mathbf{w}}\mathcal{L}(\mathbf{w}^*, \boldsymbol{\mu}^*, \boldsymbol{\lambda}^*) = 0, \quad \boldsymbol{\mu}^* \geq 0, \\ \textbf{Primal Feasibility:} & \mathbf{g}(\mathbf{w}^*) = 0, \quad \mathbf{h}(\mathbf{w}^*) \leq 0, \\ \textbf{Complementary Slackness:} & \boldsymbol{\mu}_i^* \mathbf{h}_i(\mathbf{w}^*) = 0, \quad \forall\, i\end{array}}$$

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

Optimality conditions for NLP with equality and/or inequality constraints:

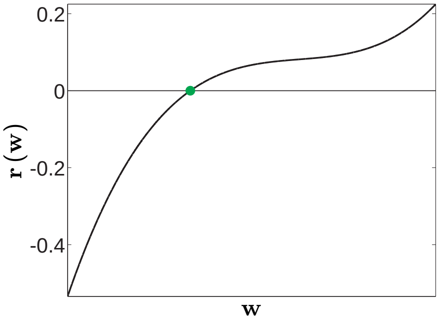
- **1st-Order Necessary Conditions**: A (local) optimum $\mathbf{w}^\star$ satisfying LICQ of a (differentiable) NLP corresponds to a unique KKT point
- **2nd-Order Sufficient Conditions** require positivity of the Hessian $\nabla^2_{\mathbf{w}}\mathcal{L}$ in all critical feasible directions at the solution

Outline

1. KKT conditions - Quick Reminder
2. The Newton method
3. Newton on the KKT conditions
4. Sequential Quadratic Programming
5. Hessian approximation
6. Maratos effect

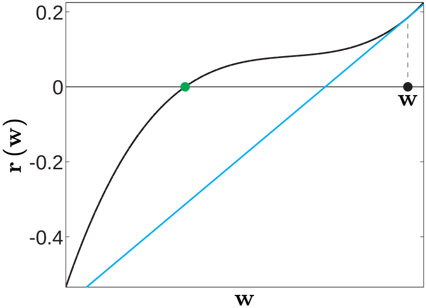
Core idea

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



Core idea

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

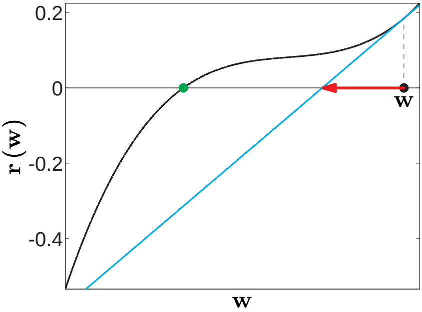


Key idea: guess $\mathbf{w}$, iterate the linear model:

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

Core idea

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

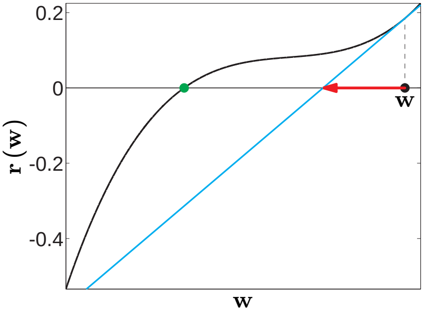


Key idea: guess $\mathbf{w}$, iterate the linear model:

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

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 $\|\mathbf{r}(\mathbf{w})\|_\infty \geq \mathrm{tol}$ **do**

Compute

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

Compute the Newton direction

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

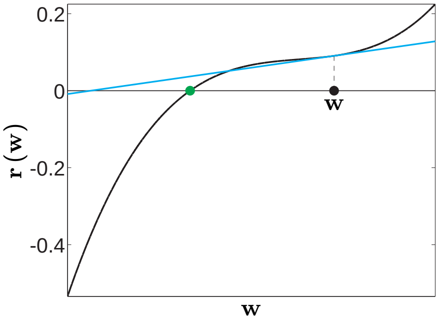
Newton step

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

return $\mathbf{w}$

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 $\|\mathbf{r}(\mathbf{w})\|_\infty \geq \mathrm{tol}$ **do**

Compute

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

Compute the Newton direction

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

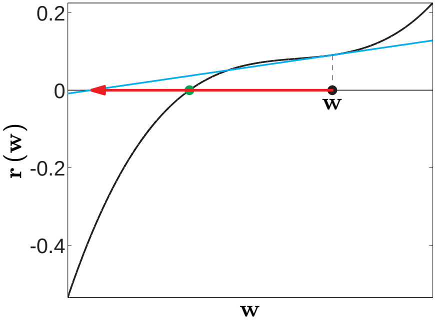
Newton step

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

return $\mathbf{w}$

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

Compute

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

Compute the Newton direction

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

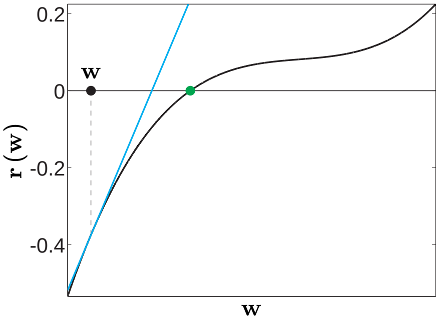
Newton step

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

return $\mathbf{w}$

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

Compute

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

Compute the Newton direction

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

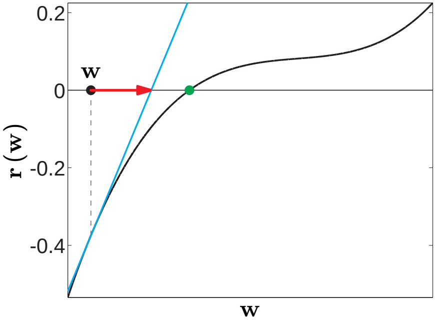
Newton step

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

return $\mathbf{w}$

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

Compute

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

Compute the Newton direction

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

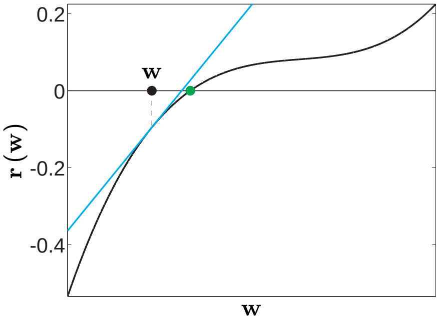
Newton step

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

return $\mathbf{w}$

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

Compute

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

Compute the Newton direction

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

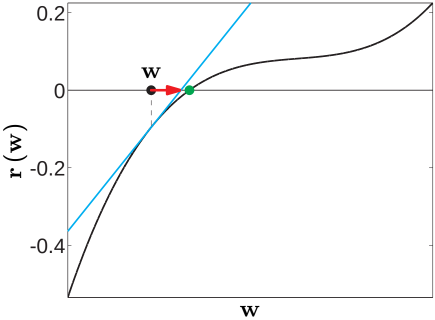
Newton step

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

return $\mathbf{w}$

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

  Compute

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

  Compute the Newton direction

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

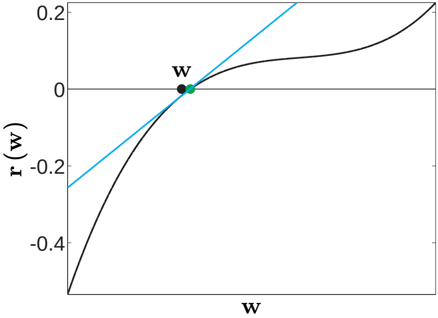
  Newton step

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

return $\mathbf{w}$

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

Compute

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

Compute the Newton direction

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

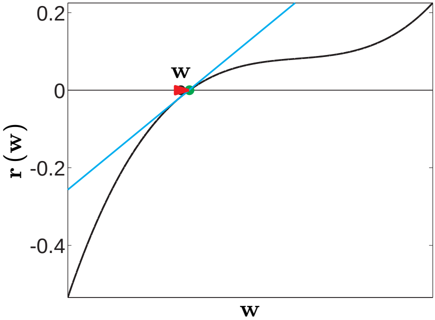
Newton step

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

return $\mathbf{w}$

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

Compute

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

Compute the Newton direction

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

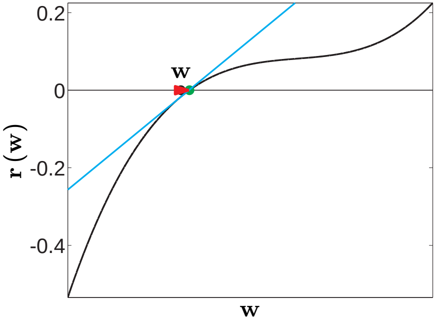
Newton step

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

return $\mathbf{w}$

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 $\|\mathbf{r}(\mathbf{w})\|_\infty \geq \text{tol}$ **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

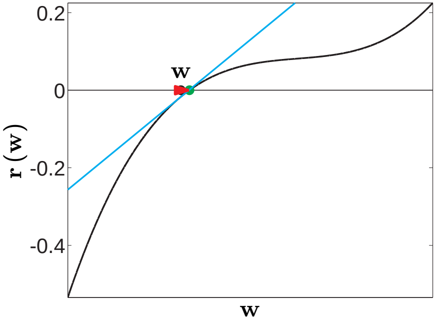
$$\mathbf{w} \leftarrow \mathbf{w} + \Delta\mathbf{w}$$

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 $\|\mathbf{r}(\mathbf{w})\|_\infty \geq \mathrm{tol}$ **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, $t \in]0, 1]$

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

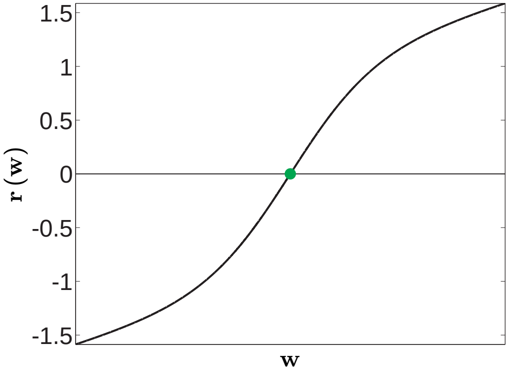
return $\mathbf{w}$

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

Why reduced steps ?

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

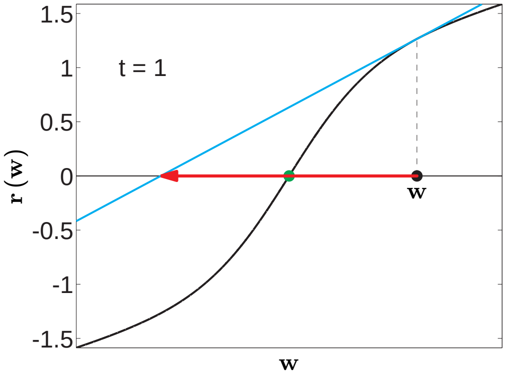
$$\nabla \mathbf{r}(\mathbf{w})^\top \Delta\mathbf{w} = -\mathbf{r}(\mathbf{w})$$
$$\mathbf{w} \leftarrow \mathbf{w} + t\Delta\mathbf{w}$$



Why reduced steps ?

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

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

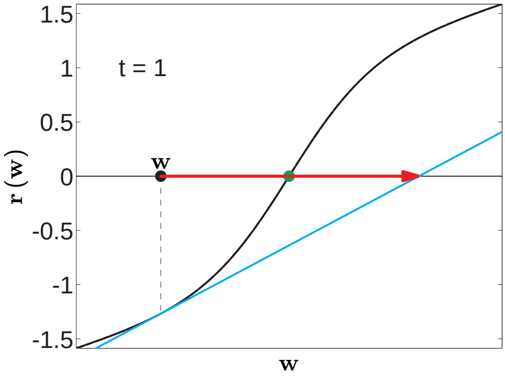


Why reduced steps ?

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

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

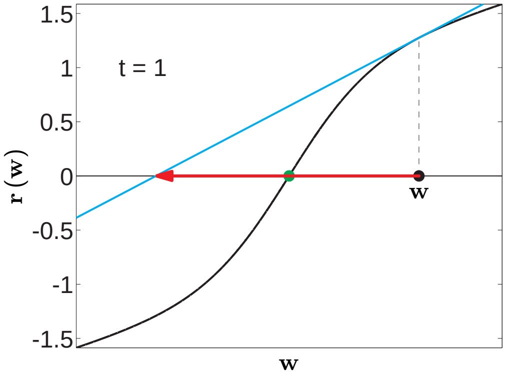
$$\mathbf{w} \leftarrow \mathbf{w} + t\Delta \mathbf{w}$$



Why reduced steps ?

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

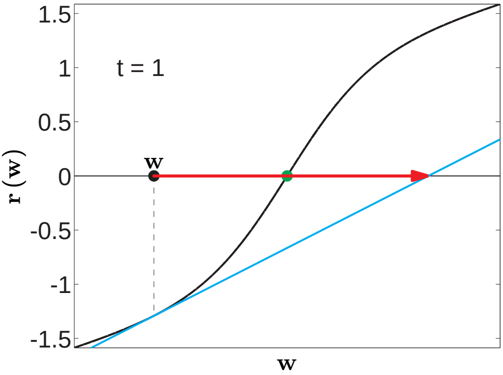
$$\nabla \mathbf{r}(\mathbf{w})^\top \Delta\mathbf{w} = -\mathbf{r}(\mathbf{w})$$
$$\mathbf{w} \leftarrow \mathbf{w} + t\Delta\mathbf{w}$$



Why reduced steps ?

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

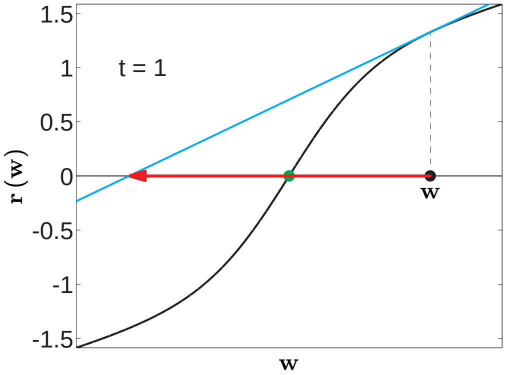
$$\nabla \mathbf{r}(\mathbf{w})^\top \Delta \mathbf{w} = -\mathbf{r}(\mathbf{w})$$
$$\mathbf{w} \leftarrow \mathbf{w} + t\Delta \mathbf{w}$$



Why reduced steps ?

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

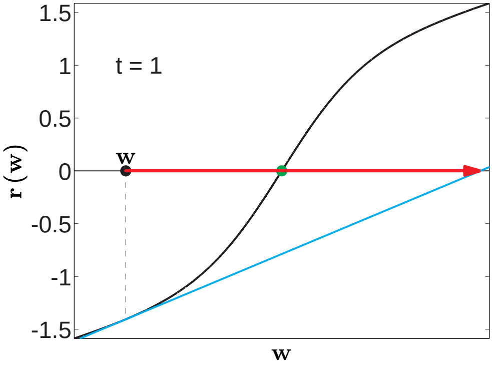
$$\nabla \mathbf{r}(\mathbf{w})^\top \Delta\mathbf{w} = -\mathbf{r}(\mathbf{w})$$
$$\mathbf{w} \leftarrow \mathbf{w} + t\Delta\mathbf{w}$$



Why reduced steps ?

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

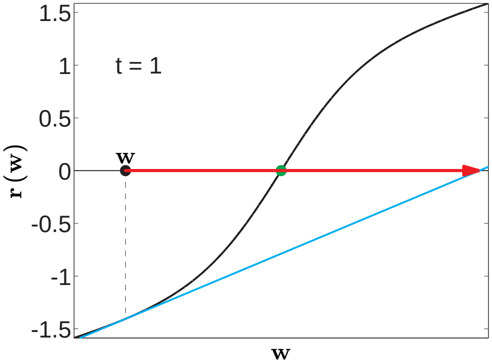
$$\nabla \mathbf{r}(\mathbf{w})^\top \Delta\mathbf{w} = -\mathbf{r}(\mathbf{w})$$
$$\mathbf{w} \leftarrow \mathbf{w} + t\Delta\mathbf{w}$$



Why reduced steps ?

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

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



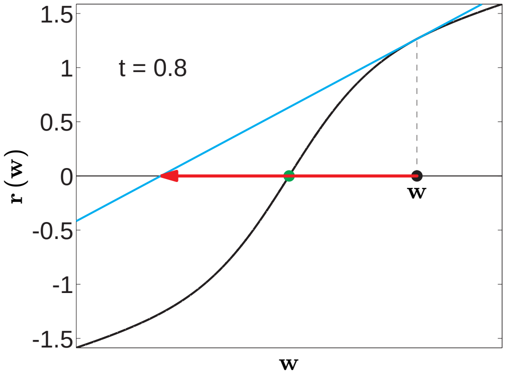
The full-step Newton iteration can be unstable !!

Why reduced steps ?

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

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

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

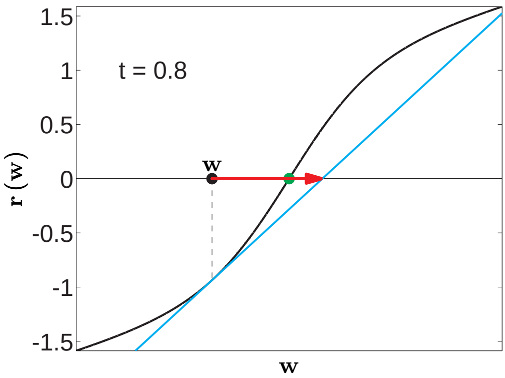


The full-step Newton iteration can be unstable !!

Why reduced steps ?

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

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

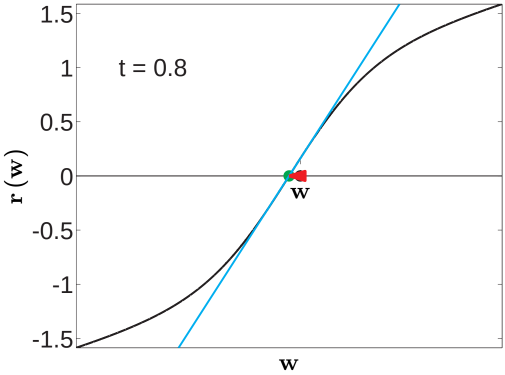


The full-step Newton iteration can be unstable !!

Why reduced steps ?

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

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

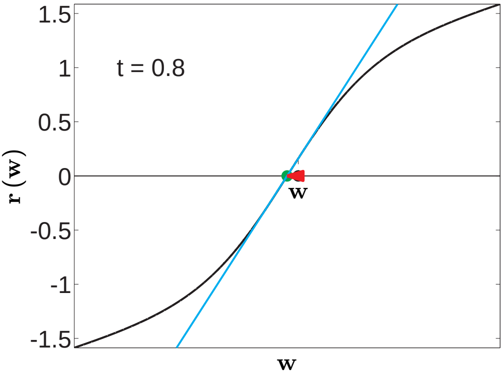


The full-step Newton iteration can be unstable !!

Why reduced steps ?

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

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



The full-step Newton iteration can be unstable !!
While the reduced-steps Newton iteration is stable...

Does Newton always work ?

Is the Newton step $\Delta \mathbf{w}$ always providing a **direction** "improving" $\mathbf{r}(\mathbf{w})$?

Does Newton always work ?

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

Does Newton always work ?

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

Does Newton always work ?

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

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

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

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

Does Newton always work ?

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

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

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

with $\|\mathbf{r}(\mathbf{w})\|^2$ 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 the Newton step $\Delta\mathbf{w}$ always providing a **direction** "improving" $\mathbf{r}(\mathbf{w})$?
I.e. is there always a $t>0$ s.t. $\|\mathbf{r}(\mathbf{w}+t\Delta\mathbf{w})\| < \|\mathbf{r}(\mathbf{w})\|$ is true ? **Yes... but**

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

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

with $\|\mathbf{r}(\mathbf{w})\|^2$ 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 the Newton step $\Delta\mathbf{w}$ always providing a **direction** "improving" $\mathbf{r}(\mathbf{w})$?
I.e. is there always a $t > 0$ s.t. $\|\mathbf{r}(\mathbf{w} + t\Delta\mathbf{w})\| < \|\mathbf{r}(\mathbf{w})\|$ is true ? **Yes... but**

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

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

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

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

Does Newton always work ?

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

How to select the step size $t \in]0,\, 1]$? Globalization...

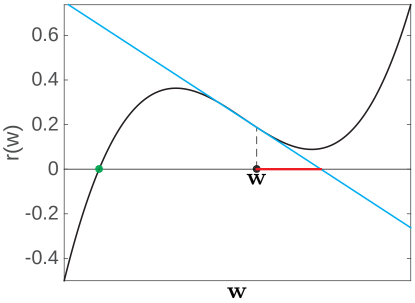
- **Line-search**: reduce t until some criteria of progression on $\|\mathbf{r}\|$ are met
- **Trust region**: confine the step $\Delta\mathbf{w}$ within a region where $\nabla\mathbf{r}(\mathbf{w})$ provides a good *model* of $\mathbf{r}(\mathbf{w})$
- **Filter techniques**: monitor progress on specific components of $\mathbf{r}(\mathbf{w})$ separately
- ...

... ensures that progress is made *in one way or another*.

Note: most of these techniques are specific to optimization.

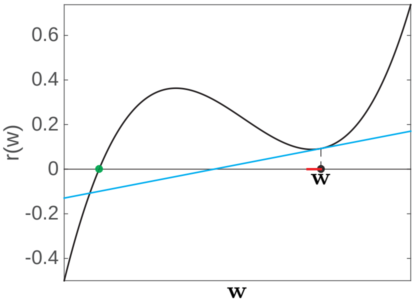
But still, Newton can fail...

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



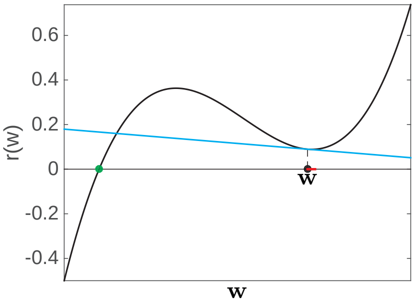
But still, Newton can fail...

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



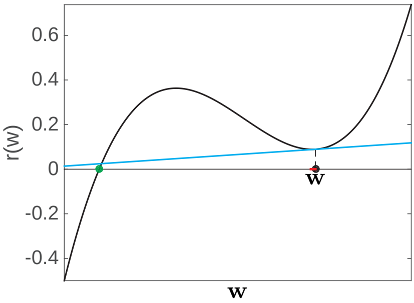
But still, Newton can fail...

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



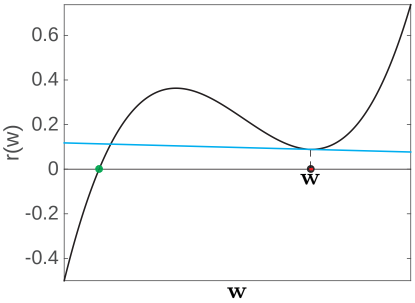
But still, Newton can fail...

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



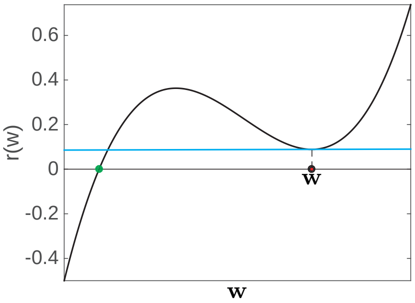
But still, Newton can fail...

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



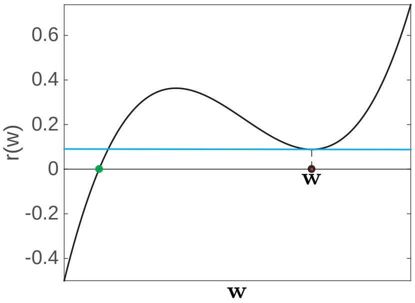
But still, Newton can fail...

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



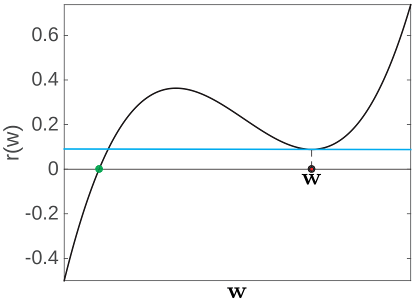
But still, Newton can fail...

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



But still, Newton can fail...

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



Newton stops with

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

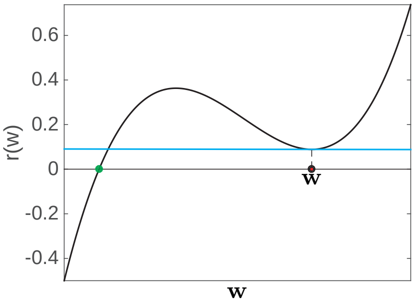
i.e. the Newton direction $\Delta \mathbf{w}$ given by

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

is undefined...

But still, Newton can fail...

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



Newton stops with

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

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

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

is undefined...

This is a common failure mode for Newton-based solvers when tackling very non-linear $\mathbf{r}$ and starting with a **poor initial guess** !!

Convergence of full-step Newton methods

Newton method:

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

Convergence of full-step Newton methods

Newton method:

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

Yields the iteration $k = 0, 1, ...$:

$$\mathbf{w}_{k+1} \leftarrow \mathbf{w}_k - \nabla \mathbf{r}(\mathbf{w}_k)^{-\top} \mathbf{r}(\mathbf{w}_k)$$

Convergence of full-step Newton methods

Newton method:

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

Yields the iteration $k = 0, 1, ...$:

$$\mathbf{w}_{k+1} \leftarrow \mathbf{w}_k - \nabla \mathbf{r}(\mathbf{w}_k)^{-\top} \mathbf{r}(\mathbf{w}_k)$$

Newton-type method (Jacobian approx.)

$$M \Delta \mathbf{w} = -\mathbf{r}(\mathbf{w})$$
$$\mathbf{w} \leftarrow \mathbf{w} + \Delta \mathbf{w}$$

Convergence of full-step Newton methods

Newton method:

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

Yields the iteration $k = 0, 1, ...$:

$$\mathbf{w}_{k+1} \leftarrow \mathbf{w}_k - \nabla \mathbf{r}(\mathbf{w}_k)^{-\top} \mathbf{r}(\mathbf{w}_k)$$

Newton-type method (Jacobian approx.)

$$M \Delta \mathbf{w} = -\mathbf{r}(\mathbf{w})$$
$$\mathbf{w} \leftarrow \mathbf{w} + \Delta \mathbf{w}$$

Yields the iteration $k = 0, 1, ...$:

$$\mathbf{w}_{k+1} \leftarrow \mathbf{w}_k - M_k^{-1} \mathbf{r}(\mathbf{w}_k)$$

Convergence of full-step Newton methods

Newton method:

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

Yields the iteration $k = 0, 1, ...$:

$$\mathbf{w}_{k+1} \leftarrow \mathbf{w}_k - \nabla \mathbf{r}(\mathbf{w}_k)^{-\top} \mathbf{r}(\mathbf{w}_k)$$

Newton-type method (Jacobian approx.)

$$M \Delta \mathbf{w} = -\mathbf{r}(\mathbf{w})$$
$$\mathbf{w} \leftarrow \mathbf{w} + \Delta \mathbf{w}$$

Yields the iteration $k = 0, 1, ...$:

$$\mathbf{w}_{k+1} \leftarrow \mathbf{w}_k - M_k^{-1} \mathbf{r}(\mathbf{w}_k)$$

Theorem: assume

- Nonlinearity of $\mathbf{r}$: $\left\| M_k^{-1} \left(\nabla \mathbf{r}(\mathbf{w})^\top - \nabla \mathbf{r}(\mathbf{w}^*)^\top \right) \right\| \leq \omega \left\| \mathbf{w} - \mathbf{w}^* \right\|$, for $\mathbf{w} \in [\mathbf{w_k}, \mathbf{w}^\star]$

Convergence of full-step Newton methods

Newton method:

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

Yields the iteration $k = 0, 1, ...$:

$$\mathbf{w}_{k+1} \leftarrow \mathbf{w}_k - \nabla \mathbf{r}(\mathbf{w}_k)^{-\top} \mathbf{r}(\mathbf{w}_k)$$

Newton-type method (Jacobian approx.)

$$M \Delta \mathbf{w} = -\mathbf{r}(\mathbf{w})$$
$$\mathbf{w} \leftarrow \mathbf{w} + \Delta \mathbf{w}$$

Yields the iteration $k = 0, 1, ...$:

$$\mathbf{w}_{k+1} \leftarrow \mathbf{w}_k - M_k^{-1} \mathbf{r}(\mathbf{w}_k)$$

Theorem: assume

- Nonlinearity of $\mathbf{r}$: $\left\| M_k^{-1} \left(\nabla \mathbf{r}(\mathbf{w})^\mathsf{T} - \nabla \mathbf{r}(\mathbf{w}^*)^\mathsf{T} \right) \right\| \leq \omega \left\| \mathbf{w} - \mathbf{w}^* \right\|$, for $\mathbf{w} \in [\mathbf{w_k}, \mathbf{w}^\star]$
- Jacobian approximation error: $\left\| M_k^{-1} (\nabla \mathbf{r}(\mathbf{w}_k)^\mathsf{T} - M_k) \right\| \leq \kappa_k < 1$

Convergence of full-step Newton methods

Newton method:

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

Yields the iteration $k = 0, 1, ...$:

$$\mathbf{w}_{k+1} \leftarrow \mathbf{w}_k - \nabla \mathbf{r}(\mathbf{w}_k)^{-\top} \mathbf{r}(\mathbf{w}_k)$$

Newton-type method (Jacobian approx.)

$$M\Delta\mathbf{w} = -\mathbf{r}(\mathbf{w})$$
$$\mathbf{w} \leftarrow \mathbf{w} + \Delta\mathbf{w}$$

Yields the iteration $k = 0, 1, ...$:

$$\mathbf{w}_{k+1} \leftarrow \mathbf{w}_k - M_k^{-1} \mathbf{r}(\mathbf{w}_k)$$

Theorem: assume

- Nonlinearity of $\mathbf{r}$: $\left\| M_k^{-1} \left(\nabla \mathbf{r}(\mathbf{w})^\mathsf{T} - \nabla \mathbf{r}(\mathbf{w}^*)^\mathsf{T} \right) \right\| \leq \omega \left\| \mathbf{w} - \mathbf{w}^* \right\|$, for $\mathbf{w} \in [\mathbf{w_k}, \mathbf{w}^\star]$
- Jacobian approximation error: $\left\| M_k^{-1} (\nabla \mathbf{r}(\mathbf{w}_k)^\mathsf{T} - M_k) \right\| \leq \kappa_k < 1$
- Good initial guess $\left\| \mathbf{w}_0 - \mathbf{w}^* \right\| \leq \frac{2}{\omega}(1 - \max\{\kappa_k\})$

Convergence of full-step Newton methods

Newton method:

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

Yields the iteration $k = 0, 1, ...$:

$$\mathbf{w}_{k+1} \leftarrow \mathbf{w}_k - \nabla \mathbf{r}(\mathbf{w}_k)^{-\top} \mathbf{r}(\mathbf{w}_k)$$

Newton-type method (Jacobian approx.)

$$M\Delta\mathbf{w} = -\mathbf{r}(\mathbf{w})$$
$$\mathbf{w} \leftarrow \mathbf{w} + \Delta\mathbf{w}$$

Yields the iteration $k = 0, 1, ...$:

$$\mathbf{w}_{k+1} \leftarrow \mathbf{w}_k - M_k^{-1} \mathbf{r}(\mathbf{w}_k)$$

Theorem: assume

- Nonlinearity of $\mathbf{r}$: $\left\| M_k^{-1} \left(\nabla \mathbf{r}(\mathbf{w})^\mathsf{T} - \nabla \mathbf{r}(\mathbf{w}^*)^\mathsf{T} \right) \right\| \leq \omega \left\| \mathbf{w} - \mathbf{w}^* \right\|$, for $\mathbf{w} \in [\mathbf{w_k}, \mathbf{w}^\star]$
- Jacobian approximation error: $\left\| M_k^{-1} (\nabla \mathbf{r}(\mathbf{w}_k)^\mathsf{T} - M_k) \right\| \leq \kappa_k < 1$
- Good initial guess $\|\mathbf{w}_0 - \mathbf{w}^*\| \leq \frac{2}{\omega}(1 - \max\{\kappa_k\})$

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

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

Convergence of full-step Newton methods

Newton method:

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

Yields the iteration $k = 0, 1, ...$:

$$\mathbf{w}_{k+1} \leftarrow \mathbf{w}_k - \nabla \mathbf{r}(\mathbf{w}_k)^{-\top} \mathbf{r}(\mathbf{w}_k)$$

Newton-type method (Jacobian approx.)

$$M\Delta\mathbf{w} = -\mathbf{r}(\mathbf{w})$$
$$\mathbf{w} \leftarrow \mathbf{w} + \Delta\mathbf{w}$$

Yields the iteration $k = 0, 1, ...$:

$$\mathbf{w}_{k+1} \leftarrow \mathbf{w}_k - M_k^{-1}\mathbf{r}(\mathbf{w}_k)$$

Theorem: assume

- Nonlinearity of $\mathbf{r}$: $\left\| M_k^{-1} \left(\nabla\mathbf{r}(\mathbf{w})^\mathsf{T} - \nabla\mathbf{r}(\mathbf{w}^*)^\mathsf{T} \right) \right\| \leq \omega \left\| \mathbf{w} - \mathbf{w}^* \right\|$, for $\mathbf{w} \in [\mathbf{w_k}, \mathbf{w}^\star]$
- Jacobian approximation error: $\left\| M_k^{-1} (\nabla\mathbf{r}(\mathbf{w}_k)^\mathsf{T} - M_k) \right\| \leq \kappa_k < 1$
- Good initial guess $\left\|\mathbf{w}_0 - \mathbf{w}^*\right\| \leq \frac{2}{\omega}(1 - \max\{\kappa_k\})$

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

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

What about reduced steps ? Slow convergence when $t < 1$ (damped phase). When full steps become feasible, fast convergence to the solution.

Newton methods - Short Survival Guide

Exact Newton method:

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

Newton-type method

$$M\Delta\mathbf{w} = -\mathbf{r}(\mathbf{w})$$
$$\mathbf{w} \leftarrow \mathbf{w} + t\Delta\mathbf{w}$$

Newton methods - Short Survival Guide

Exact Newton method:

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

Newton-type method

$$M\Delta \mathbf{w} = -\mathbf{r}(\mathbf{w})$$
$$\mathbf{w} \leftarrow \mathbf{w} + t\Delta \mathbf{w}$$

- Exact Newton direction $\Delta \mathbf{w}$ improves $\mathbf{r}$ for a sufficiently small step size $t \in]0, 1]$

Newton methods - Short Survival Guide

Exact Newton method:

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

Newton-type method

$$M\Delta\mathbf{w} = -\mathbf{r}(\mathbf{w})$$
$$\mathbf{w} \leftarrow \mathbf{w} + t\Delta\mathbf{w}$$

- Exact Newton direction $\Delta\mathbf{w}$ improves $\mathbf{r}$ for a sufficiently small step size $t \in]0, 1]$
- Inexact Newton direction $\Delta\mathbf{w}$ improves $\mathbf{r}$ for a sufficiently small step size $t \in]0, 1]$ <u>if</u> $M > 0$

Newton methods - Short Survival Guide

Exact Newton method:

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

Newton-type method

$$M\Delta\mathbf{w} = -\mathbf{r}(\mathbf{w})$$
$$\mathbf{w} \leftarrow \mathbf{w} + t\Delta\mathbf{w}$$

- Exact Newton direction $\Delta\mathbf{w}$ improves $\mathbf{r}$ for a sufficiently small step size $t \in]0,\, 1]$
- Inexact Newton direction $\Delta\mathbf{w}$ improves $\mathbf{r}$ for a sufficiently small step size $t \in]0,\, 1]$ <u>if</u> $M > 0$
- Exact full ($t = 1$) Newton steps converge quadratically if close enough to the solution

Newton methods - Short Survival Guide

Exact Newton method:

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

Newton-type method

$$M\Delta\mathbf{w} = -\mathbf{r}(\mathbf{w})$$
$$\mathbf{w} \leftarrow \mathbf{w} + t\Delta\mathbf{w}$$

- Exact Newton direction $\Delta\mathbf{w}$ improves $\mathbf{r}$ for a sufficiently small step size $t \in]0, 1]$
- Inexact Newton direction $\Delta\mathbf{w}$ improves $\mathbf{r}$ for a sufficiently small step size $t \in]0, 1]$ <u>if</u> $M > 0$
- Exact full ($t = 1$) Newton steps converge quadratically if close enough to the solution
- Inexact full ($t = 1$) Newton steps converge linearly if close enough to the solution <u>and</u> if the Jacobian approximation is "sufficiently good"

Newton methods - Short Survival Guide

Exact Newton method:

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

Newton-type method

$$M\Delta\mathbf{w} = -\mathbf{r}(\mathbf{w})$$
$$\mathbf{w} \leftarrow \mathbf{w} + t\Delta\mathbf{w}$$

- Exact Newton direction $\Delta\mathbf{w}$ improves $\mathbf{r}$ for a sufficiently small step size $t \in]0, 1]$
- Inexact Newton direction $\Delta\mathbf{w}$ improves $\mathbf{r}$ for a sufficiently small step size $t \in]0, 1]$ <u>if</u> $M > 0$
- Exact full ($t = 1$) Newton steps converge quadratically if close enough to the solution
- Inexact full ($t = 1$) Newton steps converge linearly if close enough to the solution <u>and</u> if the Jacobian approximation is "sufficiently good"
- Newton iteration fails if $\nabla\mathbf{r}$ becomes singular

Newton methods - Short Survival Guide

Exact Newton method:

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

Newton-type method

$$M\Delta\mathbf{w} = -\mathbf{r}(\mathbf{w})$$
$$\mathbf{w} \leftarrow \mathbf{w} + t\Delta\mathbf{w}$$

- Exact Newton direction $\Delta\mathbf{w}$ improves $\mathbf{r}$ for a sufficiently small step size $t \in]0, 1]$
- Inexact Newton direction $\Delta\mathbf{w}$ improves $\mathbf{r}$ for a sufficiently small step size $t \in]0, 1]$ if $M > 0$
- Exact full ($t = 1$) Newton steps converge quadratically if close enough to the solution
- Inexact full ($t = 1$) Newton steps converge linearly if close enough to the solution and if the Jacobian approximation is "sufficiently good"
- Newton iteration fails if $\nabla \mathbf{r}$ becomes singular
- Newton methods with globalization converge in two phases: damped (slow) phase when reduced steps ($t < 1$) are needed, quadratic/ linear when full steps are possible.

Outline

Core idea

A vast majority of solvers try to find a KKT point $\mathbf{w}$, $\boldsymbol{\mu}$, $\boldsymbol{\lambda}$ i.e:

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

Core idea

A vast majority of solvers try to find a KKT point $\mathbf{w}$, $\boldsymbol{\mu}$, $\boldsymbol{\lambda}$ i.e:

$$\boxed{\begin{array}{rl} \textbf{Primal Feasibility:} & \mathbf{g}(\mathbf{w}) = 0, \quad \mathbf{h}(\mathbf{w}) \leq 0, \\ \textbf{Dual Feasibility:} & \nabla_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \boldsymbol{\mu}, \boldsymbol{\lambda}) = 0, \quad \boldsymbol{\mu} \geq 0, \\ \textbf{Complementarity Slackness:} & \boldsymbol{\mu}_i \mathbf{h}_i(\mathbf{w}) = 0, \quad i = 1, \dots \end{array}}$$

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.:

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

Core idea

A vast majority of solvers try to find a KKT point $\mathbf{w}$, $\boldsymbol{\mu}$, $\boldsymbol{\lambda}$ i.e:

$$\boxed{\begin{array}{rl} \textbf{Primal Feasibility:} & \mathbf{g}(\mathbf{w}) = 0, \quad \mathbf{h}(\mathbf{w}) \leq 0, \\ \textbf{Dual Feasibility:} & \nabla_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \boldsymbol{\mu}, \boldsymbol{\lambda}) = 0, \quad \boldsymbol{\mu} \geq 0, \\ \textbf{Complementarity Slackness:} & \boldsymbol{\mu}_i \mathbf{h}_i(\mathbf{w}) = 0, \quad i = 1, \dots \end{array}}$$

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.:

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

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

Solve...

$$\mathbf{r}(\mathbf{w}, \boldsymbol{\lambda}) = \left[\begin{array}{c} \nabla_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \boldsymbol{\lambda}) \\ \mathbf{g}(\mathbf{w}) \end{array} \right] = 0$$

Core idea

A vast majority of solvers try to find a KKT point $\mathbf{w},\ \boldsymbol{\mu},\ \boldsymbol{\lambda}$ i.e:

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

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.:

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

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

$$\mathbf{r}\left(\mathbf{w},\boldsymbol{\lambda}\right) = \left[\begin{array}{c} \nabla_{\mathbf{w}}\mathcal{L}(\mathbf{w},\boldsymbol{\lambda}) \\ \mathbf{g}(\mathbf{w}) \end{array}\right] = 0$$

Newton direction

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

Newton method on the KKT conditions

KKT conditions

$$\mathbf{r}\left(\mathbf{w},\boldsymbol{\lambda}\right) = \left[\begin{array}{c} \nabla_{\mathbf{w}}\mathcal{L}(\mathbf{w},\boldsymbol{\lambda}) \\ \mathbf{g}(\mathbf{w}) \end{array}\right] = 0$$

Newton direction

$$\nabla\mathbf{r}\left(\mathbf{w},\boldsymbol{\lambda}\right)^{\mathsf{T}} \left[\begin{array}{c} \Delta\mathbf{w} \\ \Delta\boldsymbol{\lambda} \end{array}\right] = -\mathbf{r}\left(\mathbf{w},\boldsymbol{\lambda}\right)$$

Given by:

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

Newton method on the KKT conditions

KKT conditions

$$\mathbf{r}\left(\mathbf{w},\boldsymbol{\lambda}\right)=\left[\begin{array}{c}\nabla_{\mathbf{w}}\mathcal{L}(\mathbf{w},\boldsymbol{\lambda})\\ \mathbf{g}(\mathbf{w})\end{array}\right]=0$$

Newton direction

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

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

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

Newton method on the KKT conditions

KKT conditions

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

Newton direction

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

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

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

Newton method on the KKT conditions

KKT conditions

$$\mathbf{r}\left(\mathbf{w},\boldsymbol{\lambda}\right) = \left[\begin{array}{c} \nabla_{\mathbf{w}}\mathcal{L}(\mathbf{w},\boldsymbol{\lambda}) \\ \mathbf{g}(\mathbf{w}) \end{array}\right] = 0$$

Newton direction

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

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

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

Newton method on the KKT conditions

KKT conditions

$$\mathbf{r}\left(\mathbf{w},\boldsymbol{\lambda}\right) = \left[\begin{array}{c} \nabla_{\mathbf{w}}\mathcal{L}(\mathbf{w},\boldsymbol{\lambda}) \\ \mathbf{g}(\mathbf{w}) \end{array}\right] = 0$$

Newton direction

$$\nabla\mathbf{r}\left(\mathbf{w},\boldsymbol{\lambda}\right)^{\mathsf{T}}\left[\begin{array}{c} \Delta\mathbf{w} \\ \Delta\boldsymbol{\lambda} \end{array}\right] = -\mathbf{r}\left(\mathbf{w},\boldsymbol{\lambda}\right)$$

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

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

Newton method on the KKT conditions

KKT conditions

$$\mathbf{r}(\mathbf{w},\boldsymbol{\lambda}) = \left[\begin{array}{c} \nabla_{\mathbf{w}}\mathcal{L}(\mathbf{w},\boldsymbol{\lambda}) \\ \mathbf{g}(\mathbf{w}) \end{array}\right] = 0$$

Newton direction

$$\nabla\mathbf{r}(\mathbf{w},\boldsymbol{\lambda})^{\mathsf{T}} \left[\begin{array}{c} \Delta\mathbf{w} \\ \Delta\boldsymbol{\lambda} \end{array}\right] = -\mathbf{r}(\mathbf{w},\boldsymbol{\lambda})$$

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

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

The Newton direction on the KKT conditions

$$\underbrace{\left[\begin{array}{cc} \nabla^2_{\mathbf{w}}\mathcal{L}(\mathbf{w},\boldsymbol{\lambda}) & \nabla\mathbf{g}(\mathbf{w}) \\ \nabla\mathbf{g}(\mathbf{w})^{\mathsf{T}} & 0 \end{array}\right]}_{\text{KKT matrix (symmetric indefinite)}} \left[\begin{array}{c} \Delta\mathbf{w} \\ \boldsymbol{\lambda}+\Delta\boldsymbol{\lambda} \end{array}\right] = -\left[\begin{array}{c} \nabla\Phi(\mathbf{w}) \\ \mathbf{g}(\mathbf{w}) \end{array}\right]$$

Newton method on the KKT conditions

KKT conditions

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

Newton direction

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

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

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

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

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

Newton direction

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

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

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

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. Note: update of the dual variable is $\boldsymbol{\lambda}^+ = \boldsymbol{\lambda} + \Delta\boldsymbol{\lambda}$

Newton method on the KKT conditions

KKT conditions

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

Newton direction

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

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

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

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. Note: update of the dual variable is $\boldsymbol{\lambda}^{+} = \boldsymbol{\lambda} + \Delta\boldsymbol{\lambda}$

- $\nabla_{\mathbf{w}}\mathcal{L}(\mathbf{w},\boldsymbol{\lambda})$ is not needed for computing the Newton step

Newton method on the KKT conditions

KKT conditions

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

Newton direction

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

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

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

The Newton direction on the KKT conditions

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

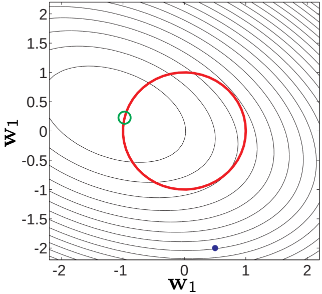
where $H(\mathbf{w},\boldsymbol{\lambda}) = \nabla^2_{\mathbf{w}}\mathcal{L}(\mathbf{w},\boldsymbol{\lambda})$ is the Hessian of the problem. Note: update of the dual variable is $\boldsymbol{\lambda}^+ = \boldsymbol{\lambda} + \Delta\boldsymbol{\lambda}$

- $\nabla_{\mathbf{w}}\mathcal{L}(\mathbf{w},\boldsymbol{\lambda})$ is not needed for computing the Newton step
- The updated dual variables $\boldsymbol{\lambda}^+$ are readily provided !

Newton Iteration for Optimization - Example

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

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



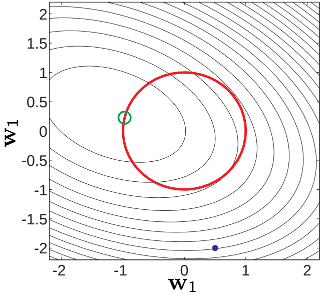
Newton Iteration for Optimization - Example

Iterate:

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

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

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



Newton Iteration for Optimization - Example

Iterate:

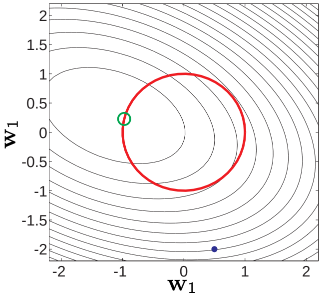
$$\begin{bmatrix} \color{blue}{H} & \nabla \mathbf{g} \\ \nabla \mathbf{g}^\top & 0 \end{bmatrix} \begin{bmatrix} \Delta \mathbf{w} \\ \boldsymbol{\lambda}^+ \end{bmatrix} = - \begin{bmatrix} \nabla \Phi \\ \mathbf{g} \end{bmatrix}$$

with:

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

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

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



Newton Iteration for Optimization - Example

Iterate:

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

with:

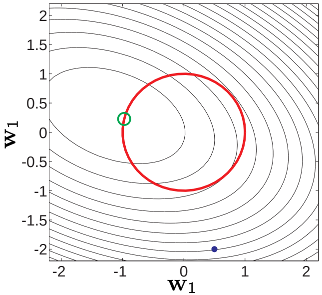
$$\nabla g(\mathbf{w}) = 2\mathbf{w} = \begin{bmatrix} 2\mathbf{w}_1 \\ 2\mathbf{w}_2 \end{bmatrix}$$

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

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

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

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



Newton Iteration for Optimization - Example

Iterate:

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

with:

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

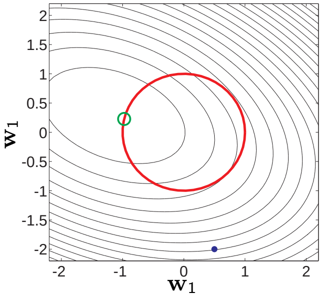
$$\mathcal{L}(\mathbf{w}, \lambda) = \Phi(\mathbf{w}) + \lambda g(\mathbf{w})$$

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

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

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

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



Newton Iteration for Optimization - Example

Iterate:

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

with:

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

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

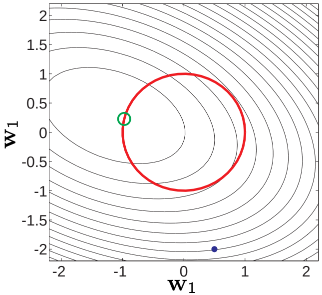
$$\nabla_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \lambda) = \begin{bmatrix} 2 & 1 \\ 1 & 4 \end{bmatrix} \mathbf{w} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 2\lambda \mathbf{w}$$

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

$$\nabla \Phi(\mathbf{w}) = \begin{bmatrix} 2\mathbf{w}_1 + \mathbf{w}_2 + 1 \\ \mathbf{w}_1 + 4\mathbf{w}_2 \end{bmatrix}$$

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

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



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

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

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

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

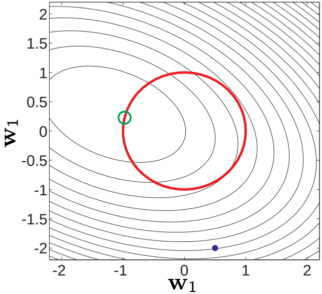
$$\mathbf{w} \leftarrow \mathbf{w} + t\Delta\mathbf{w}, \quad \boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda} + t\Delta\boldsymbol{\lambda}$$

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

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

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

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

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

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

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

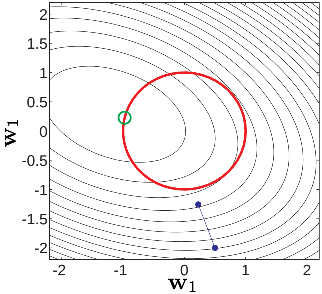
$$\mathbf{w} \leftarrow \mathbf{w} + t\Delta\mathbf{w}, \quad \boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda} + t\Delta\boldsymbol{\lambda}$$

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

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

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

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

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

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

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

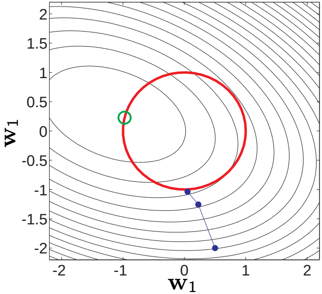
$$\mathbf{w} \leftarrow \mathbf{w} + t\Delta\mathbf{w}, \quad \boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda} + t\Delta\boldsymbol{\lambda}$$

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

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

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

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

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

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

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

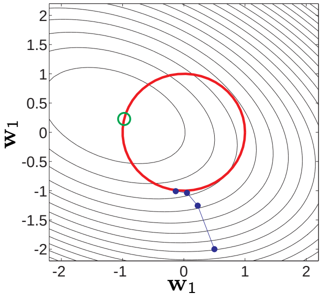
$$\mathbf{w} \leftarrow \mathbf{w} + t\Delta\mathbf{w}, \quad \boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda} + t\Delta\boldsymbol{\lambda}$$

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

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

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

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

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

Compute Newton direction

$$\left[\begin{array}{cc} H & \nabla\mathbf{g} \\ \nabla\mathbf{g}^{\top} & 0 \end{array}\right]\left[\begin{array}{c} \Delta\mathbf{w} \\ \boldsymbol{\lambda}^{+} \end{array}\right] = -\left[\begin{array}{c} \nabla\Phi \\ \mathbf{g} \end{array}\right]$$

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

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

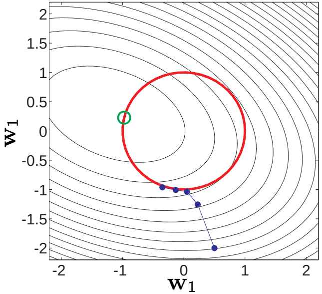
$$\mathbf{w} \leftarrow \mathbf{w} + t\Delta\mathbf{w}, \quad \boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda} + t\Delta\boldsymbol{\lambda}$$

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

$$\min_{\mathbf{w}} \ \frac{1}{2}\mathbf{w}^{\top}\left[\begin{array}{cc} 2 & 1 \\ 1 & 4 \end{array}\right]\mathbf{w} + \mathbf{w}^{\top}\left[\begin{array}{c} 1 \\ 0 \end{array}\right]$$

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

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

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

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

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

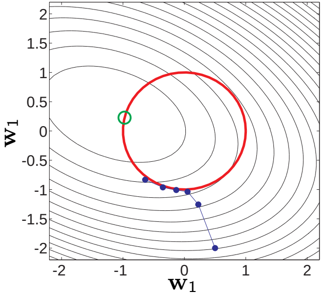
$$\mathbf{w} \leftarrow \mathbf{w} + t\Delta\mathbf{w}, \quad \boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda} + t\Delta\boldsymbol{\lambda}$$

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

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

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

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

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

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

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

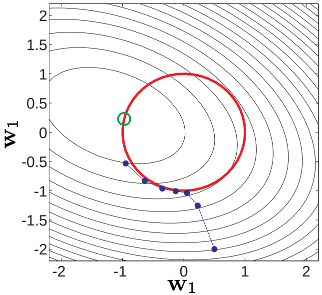
$$\mathbf{w} \leftarrow \mathbf{w} + t\Delta\mathbf{w}, \quad \boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda} + t\Delta\boldsymbol{\lambda}$$

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

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

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

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

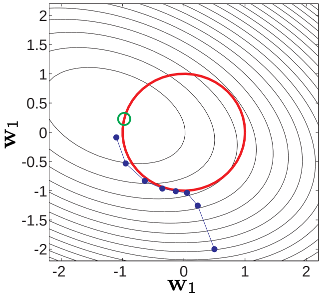
$$\mathbf{w} \leftarrow \mathbf{w} + t\Delta\mathbf{w}, \quad \boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda} + t\Delta\boldsymbol{\lambda}$$

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

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

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

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

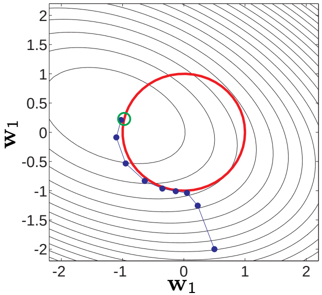
$$\mathbf{w} \leftarrow \mathbf{w} + t\Delta\mathbf{w}, \quad \boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda} + t\Delta\boldsymbol{\lambda}$$

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

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

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

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

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

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

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

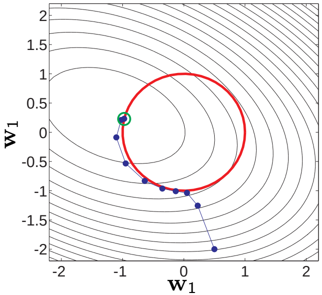
$$\mathbf{w} \leftarrow \mathbf{w} + t\Delta\mathbf{w}, \quad \boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda} + t\Delta\boldsymbol{\lambda}$$

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

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

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

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

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

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

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

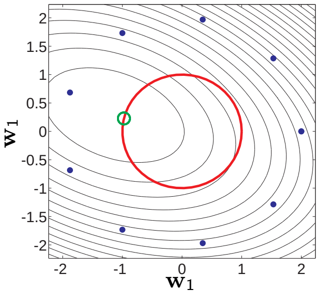
$$\mathbf{w} \leftarrow \mathbf{w} + t\Delta\mathbf{w}, \quad \boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda} + t\Delta\boldsymbol{\lambda}$$

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

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

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

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

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

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

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

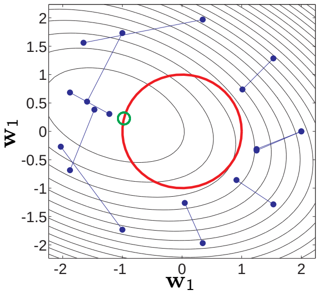
$$\mathbf{w} \leftarrow \mathbf{w} + t\Delta\mathbf{w}, \quad \boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda} + t\Delta\boldsymbol{\lambda}$$

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

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

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

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

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

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

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

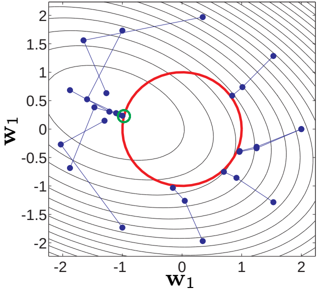
$$\mathbf{w} \leftarrow \mathbf{w} + t\Delta\mathbf{w}, \quad \boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda} + t\Delta\boldsymbol{\lambda}$$

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

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

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

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

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

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

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

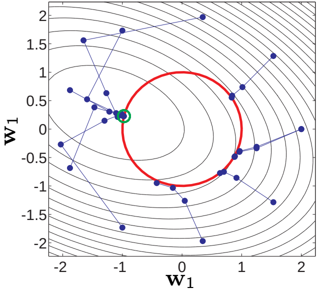
$$\mathbf{w} \leftarrow \mathbf{w} + t\Delta\mathbf{w}, \quad \boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda} + t\Delta\boldsymbol{\lambda}$$

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

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

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

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

$$\left[\begin{array}{cc} H & \nabla\mathbf{g} \\ \nabla\mathbf{g}^\top & 0 \end{array}\right]\left[\begin{array}{c} \Delta\mathbf{w} \\ \boldsymbol{\lambda}^+ \end{array}\right] = -\left[\begin{array}{c} \nabla\Phi \\ \mathbf{g} \end{array}\right]$$

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

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

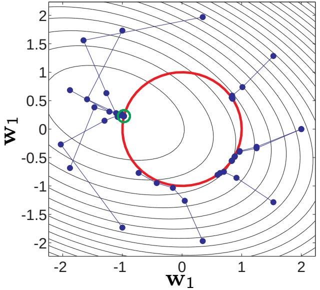
$$\mathbf{w} \leftarrow \mathbf{w} + t\Delta\mathbf{w}, \quad \boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda} + t\Delta\boldsymbol{\lambda}$$

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

$$\min_{\mathbf{w}}\ \frac{1}{2}\mathbf{w}^\top \left[\begin{array}{cc} 2 & 1 \\ 1 & 4 \end{array}\right]\mathbf{w} + \mathbf{w}^\top \left[\begin{array}{c} 1 \\ 0 \end{array}\right]$$

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

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

$$\left[\begin{array}{cc} H & \nabla\mathbf{g} \\ \nabla\mathbf{g}^\top & 0 \end{array}\right]\left[\begin{array}{c} \Delta\mathbf{w} \\ \boldsymbol{\lambda}^+ \end{array}\right] = -\left[\begin{array}{c} \nabla\Phi \\ \mathbf{g} \end{array}\right]$$

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

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

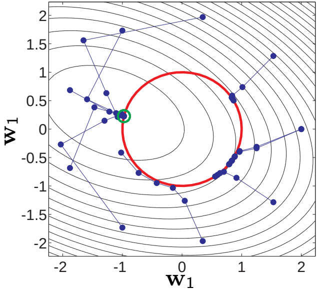
$$\mathbf{w} \leftarrow \mathbf{w} + t\Delta\mathbf{w}, \quad \boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda} + t\Delta\boldsymbol{\lambda}$$

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

$$\min_{\mathbf{w}}\ \frac{1}{2}\mathbf{w}^\top\left[\begin{array}{cc} 2 & 1 \\ 1 & 4 \end{array}\right]\mathbf{w} + \mathbf{w}^\top\left[\begin{array}{c} 1 \\ 0 \end{array}\right]$$

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

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

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

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

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

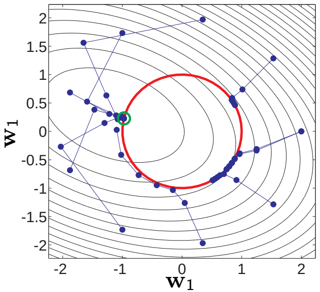
$$\mathbf{w} \leftarrow \mathbf{w} + t\Delta\mathbf{w}, \quad \boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda} + t\Delta\boldsymbol{\lambda}$$

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

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

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

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

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

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

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

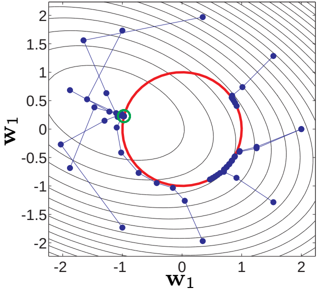
$$\mathbf{w} \leftarrow \mathbf{w} + t\Delta\mathbf{w}, \quad \boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda} + t\Delta\boldsymbol{\lambda}$$

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

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

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

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

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

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

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

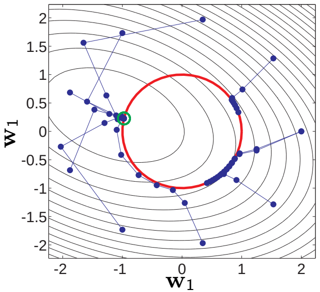
$$\mathbf{w} \leftarrow \mathbf{w} + t\Delta \mathbf{w}, \quad \boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda} + t\Delta \boldsymbol{\lambda}$$

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

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

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

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

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

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

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

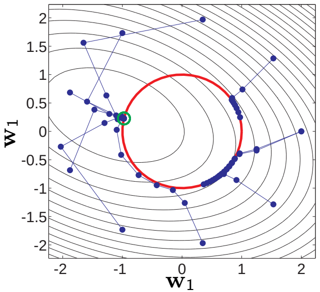
$$\mathbf{w} \leftarrow \mathbf{w} + t\Delta\mathbf{w}, \quad \boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda} + t\Delta\boldsymbol{\lambda}$$

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

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

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

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

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

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

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

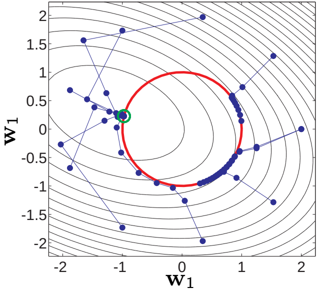
$$\mathbf{w} \leftarrow \mathbf{w} + t\Delta\mathbf{w}, \quad \boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda} + t\Delta\boldsymbol{\lambda}$$

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

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

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

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

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

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

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

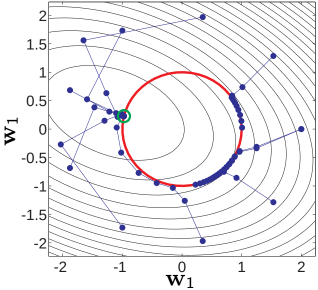
$$\mathbf{w} \leftarrow \mathbf{w} + t\Delta\mathbf{w}, \quad \boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda} + t\Delta\boldsymbol{\lambda}$$

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

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

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

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

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

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

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

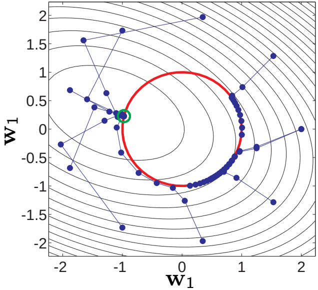
$$\mathbf{w} \leftarrow \mathbf{w} + t\Delta\mathbf{w}, \quad \boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda} + t\Delta\boldsymbol{\lambda}$$

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

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

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

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

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

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

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

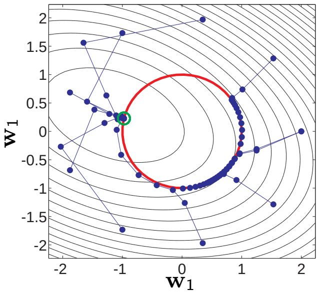
$$\mathbf{w} \leftarrow \mathbf{w} + t\Delta\mathbf{w}, \quad \boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda} + t\Delta\boldsymbol{\lambda}$$

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

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

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

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

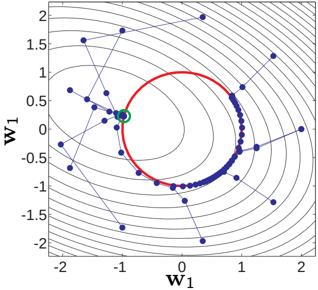
$$\mathbf{w} \leftarrow \mathbf{w} + t\Delta\mathbf{w}, \quad \boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda} + t\Delta\boldsymbol{\lambda}$$

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

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

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

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

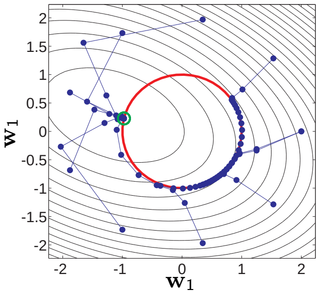
$$\mathbf{w} \leftarrow \mathbf{w} + t\Delta\mathbf{w}, \quad \boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda} + t\Delta\boldsymbol{\lambda}$$

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

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

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

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

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

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

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

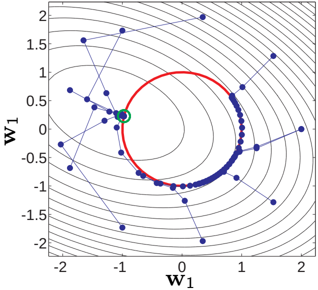
$$\mathbf{w} \leftarrow \mathbf{w} + t\Delta\mathbf{w}, \quad \boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda} + t\Delta\boldsymbol{\lambda}$$

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

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

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

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

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

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

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

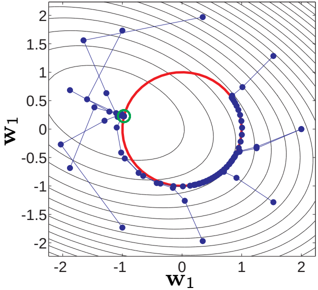
$$\mathbf{w} \leftarrow \mathbf{w} + t\Delta \mathbf{w}, \quad \boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda} + t\Delta \boldsymbol{\lambda}$$

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

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

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

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

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

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

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

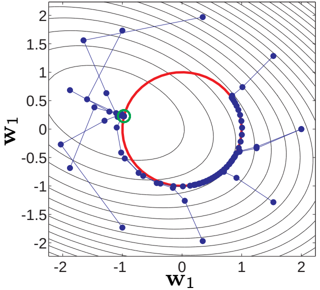
$$\mathbf{w} \leftarrow \mathbf{w} + t\Delta\mathbf{w}, \quad \boldsymbol{\lambda} \leftarrow \boldsymbol{\lambda} + t\Delta\boldsymbol{\lambda}$$

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

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

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

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



Your initial guess matters !!

Invertibility of the KKT matrix

The Newton direction of the KKT conditions

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

Invertibility of the KKT matrix

The Newton direction of the KKT conditions

$$\underbrace{\left[\begin{array}{cc} H\left(\mathbf{w},\boldsymbol{\lambda}\right) & \nabla \mathbf{g}(\mathbf{w}) \\ \nabla \mathbf{g}(\mathbf{w})^{\top} & 0 \end{array}\right]}_{\text{KKT matrix (symmetric indefinite)}} \left[\begin{array}{c} \Delta \mathbf{w} \\ \boldsymbol{\lambda}^{+} \end{array}\right] = - \left[\begin{array}{c} \nabla \Phi(\mathbf{w}) \\ \mathbf{g}(\mathbf{w}) \end{array}\right]$$

The KKT matrix is invertible if

Invertibility of the KKT matrix

The Newton direction of the KKT conditions

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

The KKT matrix is invertible if

- $\nabla \mathbf{g}(\mathbf{w})$ is full column rank (LICQ)

Invertibility of the KKT matrix

The Newton direction of the KKT conditions

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

The KKT matrix is invertible if

- $\nabla \mathbf{g}(\mathbf{w})$ is full column 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

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

The KKT matrix is invertible if

- $\nabla \mathbf{g}(\mathbf{w})$ is full column 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

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

The KKT matrix is invertible if

- $\nabla \mathbf{g}(\mathbf{w})$ is full column 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 $(\mathbf{w}, \boldsymbol{\lambda})$

If LICQ & SOSC hold at the solution, then the Newton iteration is well defined in its neighborhood

Invertibility of the KKT matrix

The Newton direction of the KKT conditions

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

The KKT matrix is invertible if

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

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

In practice, when the solution fails LICQ/SOSC, it is common to observe the solver struggling numerically, as the KKT matrix becomes increasingly ill-conditioned !!

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

If LICQ & SOSC hold at the solution, then the Newton iteration is well defined in its neighborhood

Quadratic model interpretation

Problem:

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

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

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

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

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

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

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

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

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

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

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

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

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

Dual variables $\boldsymbol{\lambda}^{+}$ given by the dual variables of the QP, i.e. $\boldsymbol{\lambda}^{+} = \boldsymbol{\lambda}_{\text{QP}}$

Proof: the KKT conditions of the QP are equivalent to the system providing the Newton direction

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

$$\left[\begin{array}{cc} H(\mathbf{w}, \boldsymbol{\lambda}) & \nabla \mathbf{g}(\mathbf{w}) \\ \nabla \mathbf{g}(\mathbf{w})^\top & 0 \end{array}\right] \left[\begin{array}{c} \Delta \mathbf{w} \\ \boldsymbol{\lambda}^+ \end{array}\right] = - \left[\begin{array}{c} \nabla \Phi(\mathbf{w}) \\ \mathbf{g}(\mathbf{w}) \end{array}\right]$$

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

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

Dual variables $\boldsymbol{\lambda}^+$ given by the dual variables of the QP, i.e. $\boldsymbol{\lambda}^+ = \boldsymbol{\lambda}_{\mathrm{QP}}$

Proof: the KKT conditions of the QP are equivalent to the system providing the Newton direction

The Newton direction is given by solving a quadratic models of the original problem !!

Outline

What about inequality constraints ?

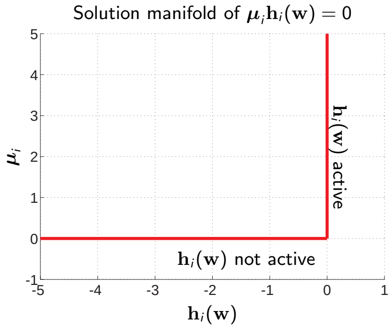
Find the "primal-dual" variables $\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\lambda}$ such that:

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:	$\boldsymbol{\mu}_i \mathbf{h}_i(\mathbf{w}) = 0, \quad i = 1, \ldots$

What about inequality constraints ?

Find the "primal-dual" variables $\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\lambda}$ such that:

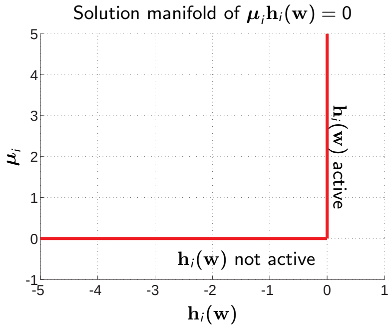
$$\boxed{\begin{array}{rl} \textbf{Primal Feasibility:} & \mathbf{g}(\mathbf{w}) = 0, \quad \mathbf{h}(\mathbf{w}) \leq 0, \\ \textbf{Dual Feasibility:} & \nabla_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \boldsymbol{\mu}, \boldsymbol{\lambda}) = 0, \quad \boldsymbol{\mu} \geq 0, \\ \textbf{Complementarity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



What about inequality constraints ?

Find the "primal-dual" variables $\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\lambda}$ such that:

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



Manifold generated by the Complementary Slackness condition is not smooth, Newton can not be used !!

Quadratic model interpretation

NLP

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

The Newton direction is given by

$$\left[\begin{array}{cc} H(\mathbf{w},\boldsymbol{\lambda}) & \nabla \mathbf{g}(\mathbf{w}) \\ \nabla \mathbf{g}(\mathbf{w})^{\top} & 0 \end{array}\right]\left[\begin{array}{c} \Delta\mathbf{w} \\ \boldsymbol{\lambda}^{+} \end{array}\right] = -\left[\begin{array}{c} \nabla\Phi(\mathbf{w}) \\ \mathbf{g}(\mathbf{w}) \end{array}\right]$$

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

Quadratic model interpretation

NLP

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

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):

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

Quadratic model interpretation

NLP

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

The Newton direction is given by

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

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):

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

Dual variables $\boldsymbol{\lambda}^+$ given by the dual variables of the QP, i.e. $\boldsymbol{\lambda}^+ = \boldsymbol{\lambda}_{\text{QP}}$

Quadratic interpretation for inequality constraints

Problem:

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

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):

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

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):

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

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.

$$\boldsymbol{\lambda}^+ = \boldsymbol{\lambda}_{\mathrm{QP}}, \quad \boldsymbol{\mu}^+ = \boldsymbol{\mu}_{\mathrm{QP}}$$

SQP Algorithm

Algorithm: SQP with line-search

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

while $\|\nabla\mathcal{L}\|_\infty$ *or* $\|\mathbf{g}\|_\infty$ *or* $\max(0, \mathbf{h}_i) \geq \mathrm{tol}$ **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

$$\begin{aligned} \min_{\Delta\mathbf{w}} \quad & \frac{1}{2}\Delta\mathbf{w}^\top H(\mathbf{w}, \boldsymbol{\lambda}, \boldsymbol{\mu})\Delta\mathbf{w} + \nabla\Phi(\mathbf{w})^\top \Delta\mathbf{w} \\ \text{s.t.} \quad & \mathbf{g}(\mathbf{w}) + \nabla\mathbf{g}(\mathbf{w})^\top \Delta\mathbf{w} = 0 \\ & \mathbf{h}(\mathbf{w}) + \nabla\mathbf{h}(\mathbf{w})^\top \Delta\mathbf{w} \leq 0 \end{aligned}$$

Select step size t to ensure progress (c.f. globalization / line-search)

Take primal step: $\mathbf{w} \leftarrow \mathbf{w} + t\Delta\mathbf{w}$

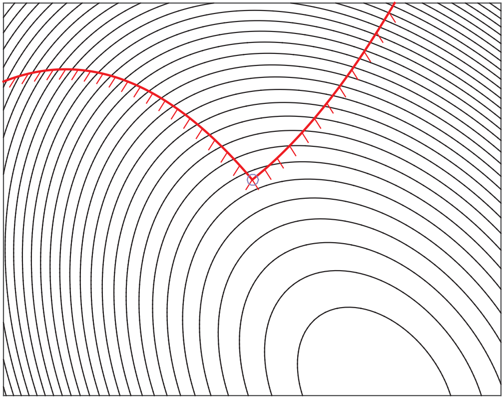
Take dual step: $\boldsymbol{\lambda} \leftarrow (1-t)\boldsymbol{\lambda} + t\boldsymbol{\lambda}_{QP}, \qquad \boldsymbol{\mu} \leftarrow (1-t)\boldsymbol{\mu} + t\boldsymbol{\mu}_{QP}$

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

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:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2}\Delta\mathbf{w}^\top H\left(\mathbf{w},\boldsymbol{\mu}\right)\Delta\mathbf{w} + \nabla\Phi\left(\mathbf{w}\right)^\top\Delta\mathbf{w} \\ \text{s.t.} \quad & \mathbf{h}\left(\mathbf{w}\right) + \nabla\mathbf{h}\left(\mathbf{w}\right)^\top\Delta\mathbf{w} \leq 0 \end{aligned}$$

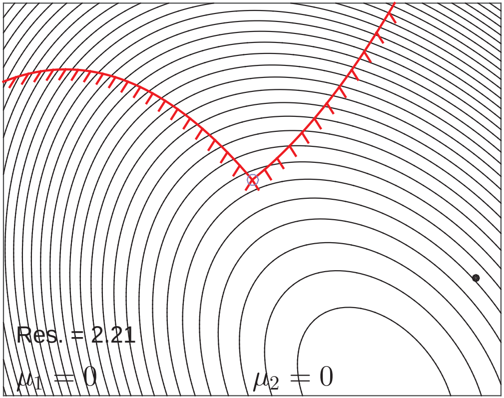
Hessian:

$$H\left(\mathbf{w},\boldsymbol{\mu}\right) = \nabla^2_{\mathbf{w}}\Phi\left(\mathbf{w}\right) + \nabla^2_{\mathbf{w}}\left(\boldsymbol{\mu}^\top\mathbf{h}\left(\mathbf{w}\right)\right)$$

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:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2}\Delta\mathbf{w}^\top H(\mathbf{w},\boldsymbol{\mu})\,\Delta\mathbf{w} + \nabla\Phi(\mathbf{w})^\top\Delta\mathbf{w} \\ \text{s.t.} \quad & \mathbf{h}(\mathbf{w}) + \nabla\mathbf{h}(\mathbf{w})^\top\Delta\mathbf{w} \leq 0 \end{aligned}$$

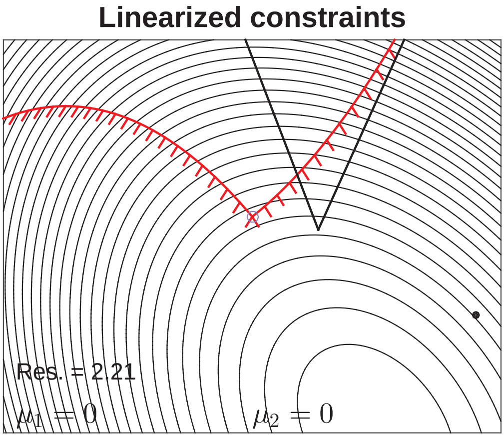
Hessian:

$$H(\mathbf{w},\boldsymbol{\mu}) = \nabla^2_{\mathbf{w}}\Phi(\mathbf{w}) + \nabla^2_{\mathbf{w}}\left(\boldsymbol{\mu}^\top\mathbf{h}(\mathbf{w})\right)$$

SQP - Illustration

NLP:

$$\min_{\mathbf{w}} \quad \frac{1}{2}\left\|\mathbf{w}-\mathbf{w}_0\right\|_Q^2$$
$$\text{s.t.} \quad \mathbf{h}(\mathbf{w}) \leq 0$$



QP:

$$\min_{\mathbf{w}} \quad \frac{1}{2}\Delta\mathbf{w}^\top H(\mathbf{w},\boldsymbol{\mu})\,\Delta\mathbf{w} + \nabla\Phi(\mathbf{w})^\top\Delta\mathbf{w}$$
$$\text{s.t.} \quad \mathbf{h}(\mathbf{w}) + \nabla\mathbf{h}(\mathbf{w})^\top\Delta\mathbf{w} \leq 0$$

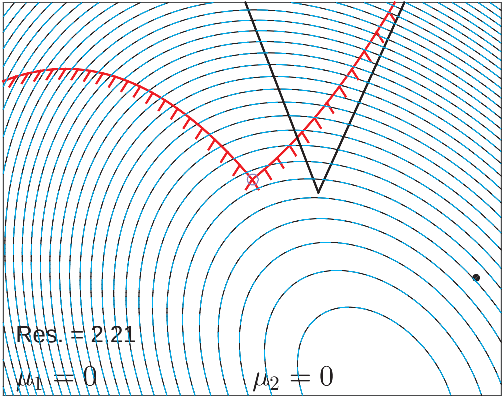
Hessian:

$$H(\mathbf{w},\boldsymbol{\mu}) = \nabla^2_{\mathbf{w}}\Phi(\mathbf{w}) + \nabla^2_{\mathbf{w}}\left(\boldsymbol{\mu}^\top\mathbf{h}(\mathbf{w})\right)$$

SQP - Illustration

NLP:

$$\min_{\mathbf{w}} \quad \frac{1}{2}\left\|\mathbf{w}-\mathbf{w}_0\right\|_Q^2$$
$$\text{s.t.} \quad \mathbf{h}(\mathbf{w}) \leq 0$$



QP:

$$\min_{\mathbf{w}} \quad \frac{1}{2}\Delta\mathbf{w}^\top H(\mathbf{w},\boldsymbol{\mu})\,\Delta\mathbf{w} + \nabla\Phi(\mathbf{w})^\top \Delta\mathbf{w}$$
$$\text{s.t.} \quad \mathbf{h}(\mathbf{w}) + \nabla\mathbf{h}(\mathbf{w})^\top \Delta\mathbf{w} \leq 0$$

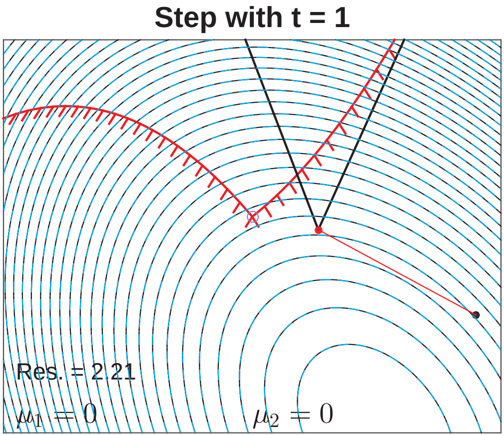
Hessian:

$$H(\mathbf{w},\boldsymbol{\mu}) = \nabla^2_{\mathbf{w}}\Phi(\mathbf{w}) + \nabla^2_{\mathbf{w}}\left(\boldsymbol{\mu}^\top\mathbf{h}(\mathbf{w})\right)$$

SQP - Illustration

NLP:

$$\min_{\mathbf{w}} \quad \frac{1}{2}\left\|\mathbf{w}-\mathbf{w}_0\right\|_Q^2$$
$$\text{s.t.} \quad \mathbf{h}(\mathbf{w}) \leq 0$$



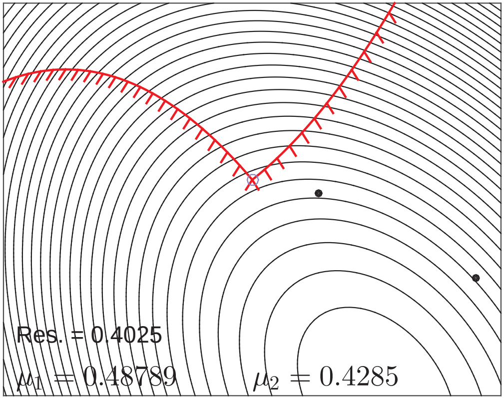
QP:

$$\min_{\mathbf{w}} \quad \frac{1}{2}\Delta\mathbf{w}^\top H(\mathbf{w},\boldsymbol{\mu})\,\Delta\mathbf{w} + \nabla\Phi(\mathbf{w})^\top\Delta\mathbf{w}$$
$$\text{s.t.} \quad \mathbf{h}(\mathbf{w}) + \nabla\mathbf{h}(\mathbf{w})^\top\Delta\mathbf{w} \leq 0$$

Hessian:

$$H(\mathbf{w},\boldsymbol{\mu}) = \nabla^2_{\mathbf{w}}\Phi(\mathbf{w}) + \nabla^2_{\mathbf{w}}\left(\boldsymbol{\mu}^\top\mathbf{h}(\mathbf{w})\right)$$

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:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2}\Delta\mathbf{w}^\top H(\mathbf{w},\boldsymbol{\mu})\Delta\mathbf{w} + \nabla\Phi(\mathbf{w})^\top \Delta\mathbf{w} \\ \text{s.t.} \quad & \mathbf{h}(\mathbf{w}) + \nabla\mathbf{h}(\mathbf{w})^\top \Delta\mathbf{w} \leq 0 \end{aligned}$$

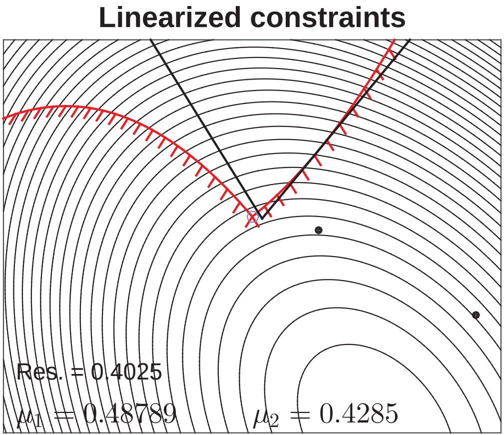
Hessian:

$$H(\mathbf{w},\boldsymbol{\mu}) = \nabla^2_{\mathbf{w}}\Phi(\mathbf{w}) + \nabla^2_{\mathbf{w}}\left(\boldsymbol{\mu}^\top\mathbf{h}(\mathbf{w})\right)$$

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:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2}\Delta\mathbf{w}^\top H(\mathbf{w},\boldsymbol{\mu})\,\Delta\mathbf{w} + \nabla\Phi(\mathbf{w})^\top \Delta\mathbf{w} \\ \text{s.t.} \quad & \mathbf{h}(\mathbf{w}) + \nabla\mathbf{h}(\mathbf{w})^\top \Delta\mathbf{w} \leq 0 \end{aligned}$$

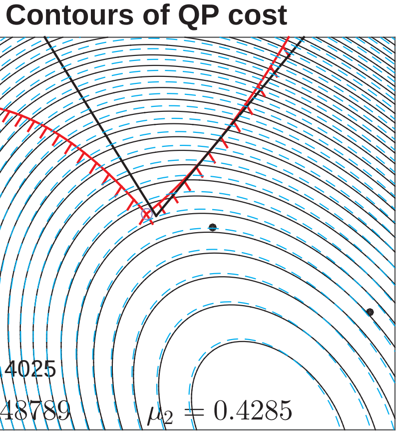
Hessian:

$$H(\mathbf{w},\boldsymbol{\mu}) = \nabla^2_{\mathbf{w}}\Phi(\mathbf{w}) + \nabla^2_{\mathbf{w}}\left(\boldsymbol{\mu}^\top\mathbf{h}(\mathbf{w})\right)$$

SQP - Illustration

NLP:

$$\min_{\mathbf{w}} \quad \frac{1}{2}\left\|\mathbf{w}-\mathbf{w}_0\right\|_Q^2$$
$$\text{s.t.} \quad \mathbf{h}(\mathbf{w}) \leq 0$$



QP:

$$\min_{\mathbf{w}} \quad \frac{1}{2}\Delta\mathbf{w}^\top H(\mathbf{w},\boldsymbol{\mu})\,\Delta\mathbf{w} + \nabla\Phi(\mathbf{w})^\top\Delta\mathbf{w}$$
$$\text{s.t.} \quad \mathbf{h}(\mathbf{w}) + \nabla\mathbf{h}(\mathbf{w})^\top\Delta\mathbf{w} \leq 0$$

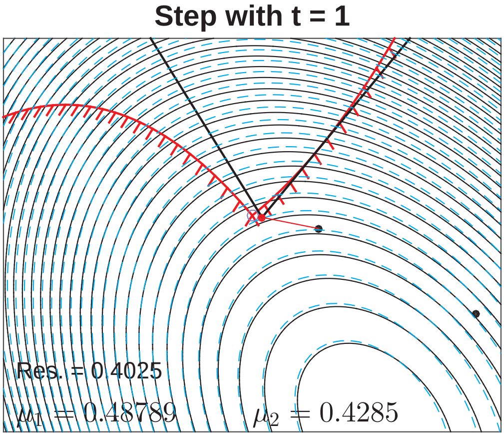
Hessian:

$$H(\mathbf{w},\boldsymbol{\mu}) = \nabla^2_{\mathbf{w}}\Phi(\mathbf{w}) + \nabla^2_{\mathbf{w}}\left(\boldsymbol{\mu}^\top\mathbf{h}(\mathbf{w})\right)$$

SQP - Illustration

NLP:

$$\min_{\mathbf{w}} \quad \frac{1}{2}\|\mathbf{w}-\mathbf{w}_0\|_Q^2$$
$$\text{s.t.} \quad \mathbf{h}(\mathbf{w}) \leq 0$$



QP:

$$\min_{\mathbf{w}} \quad \frac{1}{2}\Delta\mathbf{w}^\top H(\mathbf{w},\boldsymbol{\mu})\,\Delta\mathbf{w} + \nabla\Phi(\mathbf{w})^\top \Delta\mathbf{w}$$
$$\text{s.t.} \quad \mathbf{h}(\mathbf{w}) + \nabla\mathbf{h}(\mathbf{w})^\top \Delta\mathbf{w} \leq 0$$

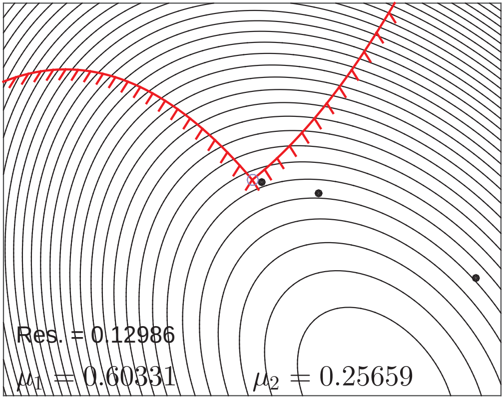
Hessian:

$$H(\mathbf{w},\boldsymbol{\mu}) = \nabla^2_{\mathbf{w}}\Phi(\mathbf{w}) + \nabla^2_{\mathbf{w}}\left(\boldsymbol{\mu}^\top \mathbf{h}(\mathbf{w})\right)$$

SQP - Illustration

NLP:

$$\begin{array}{ll} \min\limits_{\mathbf{w}} & \frac{1}{2}\left\|\mathbf{w}-\mathbf{w}_0\right\|_Q^2 \\ \text{s.t.} & \mathbf{h}(\mathbf{w}) \leq 0 \end{array}$$



QP:

$$\begin{array}{ll} \min\limits_{\mathbf{w}} & \frac{1}{2}\Delta\mathbf{w}^\top H(\mathbf{w},\boldsymbol{\mu})\Delta\mathbf{w} + \nabla\Phi(\mathbf{w})^\top\Delta\mathbf{w} \\ \text{s.t.} & \mathbf{h}(\mathbf{w}) + \nabla\mathbf{h}(\mathbf{w})^\top\Delta\mathbf{w} \leq 0 \end{array}$$

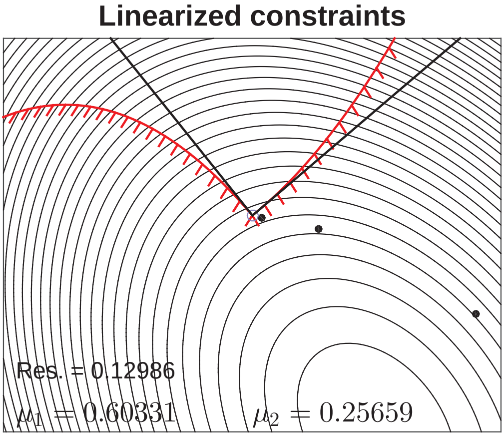
Hessian:

$$H(\mathbf{w},\boldsymbol{\mu}) = \nabla^2_{\mathbf{w}}\Phi(\mathbf{w}) + \nabla^2_{\mathbf{w}}\left(\boldsymbol{\mu}^\top\mathbf{h}(\mathbf{w})\right)$$

SQP - Illustration

NLP:

$$\min_{\mathbf{w}} \quad \frac{1}{2}\|\mathbf{w}-\mathbf{w}_0\|_Q^2$$
$$\text{s.t.} \quad \mathbf{h}(\mathbf{w}) \leq 0$$



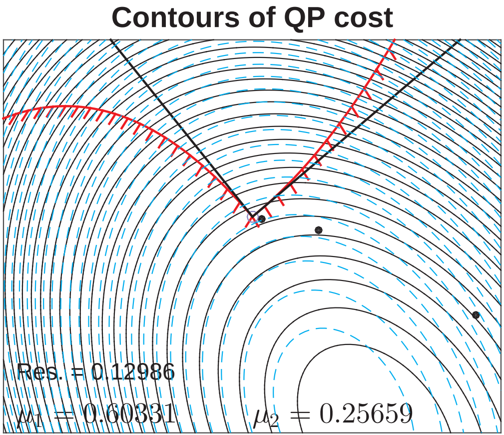
QP:

$$\min_{\mathbf{w}} \quad \frac{1}{2}\Delta\mathbf{w}^\top H(\mathbf{w},\boldsymbol{\mu})\,\Delta\mathbf{w} + \nabla\Phi(\mathbf{w})^\top\Delta\mathbf{w}$$
$$\text{s.t.} \quad \mathbf{h}(\mathbf{w}) + \nabla\mathbf{h}(\mathbf{w})^\top\Delta\mathbf{w} \leq 0$$

Hessian:

$$H(\mathbf{w},\boldsymbol{\mu}) = \nabla^2_{\mathbf{w}}\Phi(\mathbf{w}) + \nabla^2_{\mathbf{w}}\left(\boldsymbol{\mu}^\top\mathbf{h}(\mathbf{w})\right)$$

SQP - Illustration



NLP:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2}\|\mathbf{w}-\mathbf{w}_0\|_Q^2 \\ \text{s.t.} \quad & \mathbf{h}(\mathbf{w}) \leq 0 \end{aligned}$$

QP:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2}\Delta\mathbf{w}^\top H(\mathbf{w},\boldsymbol{\mu})\,\Delta\mathbf{w} + \nabla\Phi(\mathbf{w})^\top \Delta\mathbf{w} \\ \text{s.t.} \quad & \mathbf{h}(\mathbf{w}) + \nabla\mathbf{h}(\mathbf{w})^\top \Delta\mathbf{w} \leq 0 \end{aligned}$$

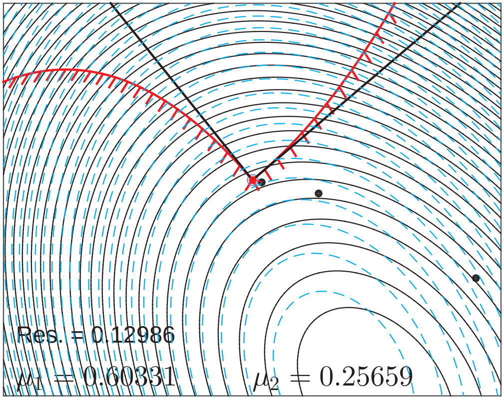
Hessian:

$$H(\mathbf{w},\boldsymbol{\mu}) = \nabla^2_{\mathbf{w}}\Phi(\mathbf{w}) + \nabla^2_{\mathbf{w}}\left(\boldsymbol{\mu}^\top\mathbf{h}(\mathbf{w})\right)$$

SQP - Illustration

NLP:

$$\min_{\mathbf{w}} \quad \frac{1}{2}\|\mathbf{w}-\mathbf{w}_0\|_Q^2$$
$$\text{s.t.} \quad \mathbf{h}(\mathbf{w}) \leq 0$$



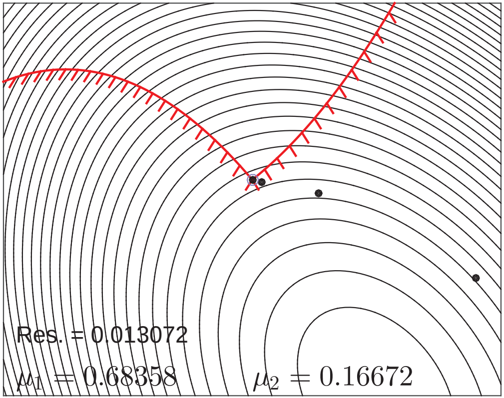
QP:

$$\min_{\mathbf{w}} \quad \frac{1}{2}\Delta\mathbf{w}^\top H(\mathbf{w},\boldsymbol{\mu})\,\Delta\mathbf{w} + \nabla\Phi(\mathbf{w})^\top\Delta\mathbf{w}$$
$$\text{s.t.} \quad \mathbf{h}(\mathbf{w}) + \nabla\mathbf{h}(\mathbf{w})^\top\Delta\mathbf{w} \leq 0$$

Hessian:

$$H(\mathbf{w},\boldsymbol{\mu}) = \nabla^2_{\mathbf{w}}\Phi(\mathbf{w}) + \nabla^2_{\mathbf{w}}\left(\boldsymbol{\mu}^\top\mathbf{h}(\mathbf{w})\right)$$

SQP - Illustration



NLP:

$$\min_{\mathbf{w}} \quad \frac{1}{2}\left\|\mathbf{w}-\mathbf{w}_0\right\|_Q^2$$
$$\text{s.t.} \quad \mathbf{h}(\mathbf{w}) \leq 0$$

QP:

$$\min_{\mathbf{w}} \quad \frac{1}{2}\Delta\mathbf{w}^\top H(\mathbf{w},\boldsymbol{\mu})\Delta\mathbf{w} + \nabla\Phi(\mathbf{w})^\top\Delta\mathbf{w}$$
$$\text{s.t.} \quad \mathbf{h}(\mathbf{w}) + \nabla\mathbf{h}(\mathbf{w})^\top\Delta\mathbf{w} \leq 0$$

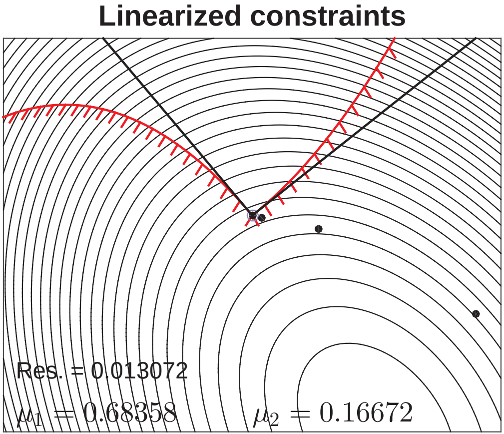
Hessian:

$$H(\mathbf{w},\boldsymbol{\mu}) = \nabla^2_{\mathbf{w}}\Phi(\mathbf{w}) + \nabla^2_{\mathbf{w}}\left(\boldsymbol{\mu}^\top\mathbf{h}(\mathbf{w})\right)$$

SQP - Illustration

NLP:

$$\begin{array}{ll} \min\limits_{\mathbf{w}} & \frac{1}{2}\left\|\mathbf{w}-\mathbf{w}_0\right\|_Q^2 \\ \text{s.t.} & \mathbf{h}(\mathbf{w}) \leq 0 \end{array}$$



QP:

$$\begin{array}{ll} \min\limits_{\mathbf{w}} & \frac{1}{2}\Delta\mathbf{w}^\top H(\mathbf{w},\boldsymbol{\mu})\,\Delta\mathbf{w} + \nabla\Phi(\mathbf{w})^\top\Delta\mathbf{w} \\ \text{s.t.} & \mathbf{h}(\mathbf{w}) + \nabla\mathbf{h}(\mathbf{w})^\top\Delta\mathbf{w} \leq 0 \end{array}$$

Hessian:

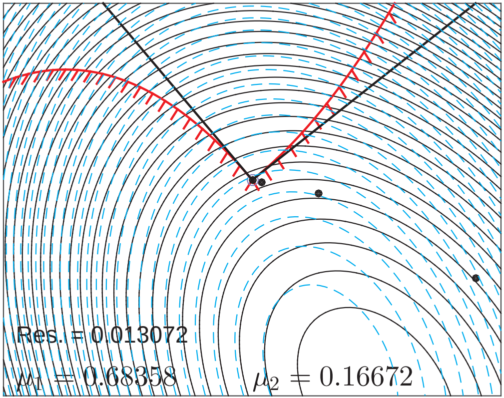
$$H(\mathbf{w},\boldsymbol{\mu}) = \nabla^2_{\mathbf{w}}\Phi(\mathbf{w}) + \nabla^2_{\mathbf{w}}\left(\boldsymbol{\mu}^\top\mathbf{h}(\mathbf{w})\right)$$

SQP - Illustration

NLP:

$$\min_{\mathbf{w}} \quad \frac{1}{2}\left\|\mathbf{w}-\mathbf{w}_0\right\|_Q^2$$
$$\text{s.t.} \quad \mathbf{h}(\mathbf{w}) \leq 0$$

Contours of QP cost



QP:

$$\min_{\mathbf{w}} \quad \frac{1}{2}\Delta\mathbf{w}^\top H(\mathbf{w},\boldsymbol{\mu})\,\Delta\mathbf{w} + \nabla\Phi(\mathbf{w})^\top \Delta\mathbf{w}$$
$$\text{s.t.} \quad \mathbf{h}(\mathbf{w}) + \nabla\mathbf{h}(\mathbf{w})^\top \Delta\mathbf{w} \leq 0$$

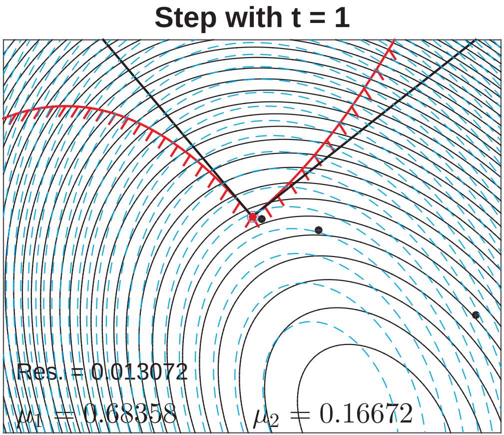
Hessian:

$$H(\mathbf{w},\boldsymbol{\mu}) = \nabla^2_{\mathbf{w}}\Phi(\mathbf{w}) + \nabla^2_{\mathbf{w}}\left(\boldsymbol{\mu}^\top\mathbf{h}(\mathbf{w})\right)$$

SQP - Illustration

NLP:

$$\min_{\mathbf{w}} \quad \frac{1}{2}\left\|\mathbf{w}-\mathbf{w}_0\right\|_Q^2$$
$$\text{s.t.} \quad \mathbf{h}(\mathbf{w}) \leq 0$$



QP:

$$\min_{\mathbf{w}} \quad \frac{1}{2}\Delta\mathbf{w}^\top H(\mathbf{w},\boldsymbol{\mu})\Delta\mathbf{w} + \nabla\Phi(\mathbf{w})^\top\Delta\mathbf{w}$$
$$\text{s.t.} \quad \mathbf{h}(\mathbf{w}) + \nabla\mathbf{h}(\mathbf{w})^\top\Delta\mathbf{w} \leq 0$$

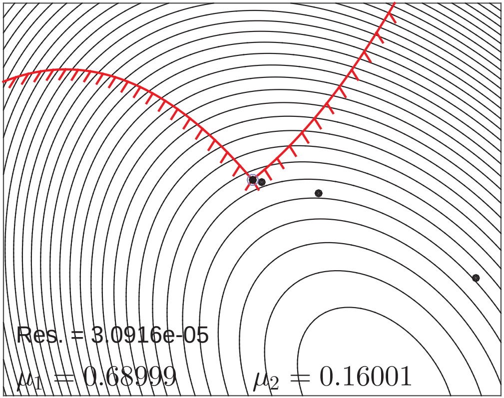
Hessian:

$$H(\mathbf{w},\boldsymbol{\mu}) = \nabla^2_{\mathbf{w}}\Phi(\mathbf{w}) + \nabla^2_{\mathbf{w}}\left(\boldsymbol{\mu}^\top\mathbf{h}(\mathbf{w})\right)$$

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:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2}\Delta\mathbf{w}^\top H(\mathbf{w},\boldsymbol{\mu})\,\Delta\mathbf{w} + \nabla\Phi(\mathbf{w})^\top\Delta\mathbf{w} \\ \text{s.t.} \quad & \mathbf{h}(\mathbf{w}) + \nabla\mathbf{h}(\mathbf{w})^\top\Delta\mathbf{w} \leq 0 \end{aligned}$$

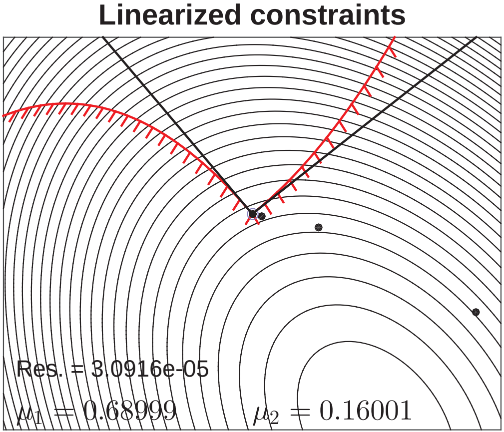
Hessian:

$$H(\mathbf{w},\boldsymbol{\mu}) = \nabla^2_{\mathbf{w}}\Phi(\mathbf{w}) + \nabla^2_{\mathbf{w}}\left(\boldsymbol{\mu}^\top\mathbf{h}(\mathbf{w})\right)$$

SQP - Illustration

NLP:

$$\min_{\mathbf{w}} \quad \frac{1}{2}\left\|\mathbf{w}-\mathbf{w}_0\right\|_Q^2$$
$$\text{s.t.} \quad \mathbf{h}(\mathbf{w}) \leq 0$$



QP:

$$\min_{\mathbf{w}} \quad \frac{1}{2}\Delta\mathbf{w}^\top H(\mathbf{w},\boldsymbol{\mu})\,\Delta\mathbf{w} + \nabla\Phi(\mathbf{w})^\top \Delta\mathbf{w}$$
$$\text{s.t.} \quad \mathbf{h}(\mathbf{w}) + \nabla\mathbf{h}(\mathbf{w})^\top \Delta\mathbf{w} \leq 0$$

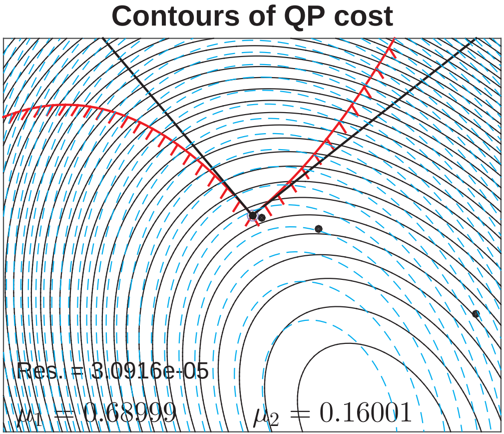
Hessian:

$$H(\mathbf{w},\boldsymbol{\mu}) = \nabla^2_{\mathbf{w}}\Phi(\mathbf{w}) + \nabla^2_{\mathbf{w}}\left(\boldsymbol{\mu}^\top\mathbf{h}(\mathbf{w})\right)$$

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:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2} \Delta \mathbf{w}^\top H(\mathbf{w}, \boldsymbol{\mu}) \Delta \mathbf{w} + \nabla \Phi(\mathbf{w})^\top \Delta \mathbf{w} \\ \text{s.t.} \quad & \mathbf{h}(\mathbf{w}) + \nabla \mathbf{h}(\mathbf{w})^\top \Delta \mathbf{w} \leq 0 \end{aligned}$$

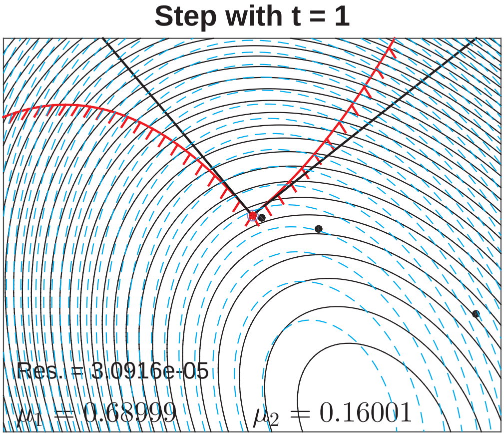
Hessian:

$$H(\mathbf{w}, \boldsymbol{\mu}) = \nabla^2_{\mathbf{w}} \Phi(\mathbf{w}) + \nabla^2_{\mathbf{w}} \left(\boldsymbol{\mu}^\top \mathbf{h}(\mathbf{w}) \right)$$

SQP - Illustration

NLP:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2}\left\|\mathbf{w}-\mathbf{w}_0\right\|_Q^2 \\ \text{s.t.} \quad & \mathbf{h}(\mathbf{w}) \leq 0 \end{aligned}$$



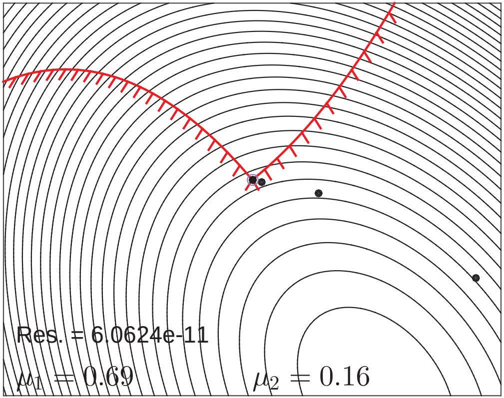
QP:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2}\Delta\mathbf{w}^\top H(\mathbf{w},\boldsymbol{\mu})\,\Delta\mathbf{w} + \nabla\Phi(\mathbf{w})^\top \Delta\mathbf{w} \\ \text{s.t.} \quad & \mathbf{h}(\mathbf{w}) + \nabla\mathbf{h}(\mathbf{w})^\top \Delta\mathbf{w} \leq 0 \end{aligned}$$

Hessian:

$$H(\mathbf{w},\boldsymbol{\mu}) = \nabla^2_{\mathbf{w}}\Phi(\mathbf{w}) + \nabla^2_{\mathbf{w}}\left(\boldsymbol{\mu}^\top\mathbf{h}(\mathbf{w})\right)$$

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:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \frac{1}{2}\Delta\mathbf{w}^\top H(\mathbf{w},\boldsymbol{\mu})\,\Delta\mathbf{w} + \nabla\Phi(\mathbf{w})^\top \Delta\mathbf{w} \\ \text{s.t.} \quad & \mathbf{h}(\mathbf{w}) + \nabla\mathbf{h}(\mathbf{w})^\top \Delta\mathbf{w} \leq 0 \end{aligned}$$

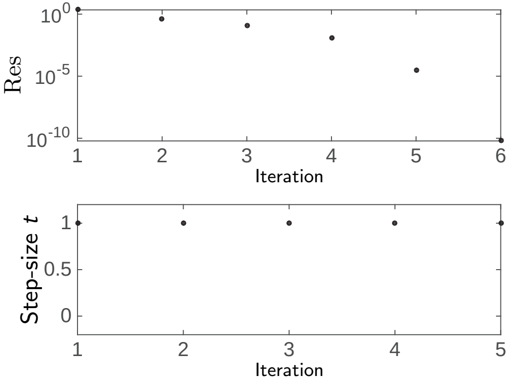
Hessian:

$$H(\mathbf{w},\boldsymbol{\mu}) = \nabla^2_{\mathbf{w}}\Phi(\mathbf{w}) + \nabla^2_{\mathbf{w}}\left(\boldsymbol{\mu}^\top\mathbf{h}(\mathbf{w})\right)$$

SQP - Illustration

NLP:

$$\begin{array}{ll} \min\limits_{\mathbf{w}} & \frac{1}{2}\left\|\mathbf{w}-\mathbf{w}_0\right\|_Q^2 \\ \text{s.t.} & \mathbf{h}\left(\mathbf{w}\right) \leq 0 \end{array}$$



QP:

$$\begin{array}{ll} \min\limits_{\mathbf{w}} & \frac{1}{2}\Delta\mathbf{w}^\top H\left(\mathbf{w},\boldsymbol{\mu}\right)\Delta\mathbf{w} + \nabla\Phi\left(\mathbf{w}\right)^\top\Delta\mathbf{w} \\ \text{s.t.} & \mathbf{h}\left(\mathbf{w}\right) + \nabla\mathbf{h}\left(\mathbf{w}\right)^\top\Delta\mathbf{w} \leq 0 \end{array}$$

Hessian:

$$H\left(\mathbf{w},\boldsymbol{\mu}\right) = \nabla^2_{\mathbf{w}}\Phi\left(\mathbf{w}\right) + \nabla^2_{\mathbf{w}}\left(\boldsymbol{\mu}^\top\mathbf{h}\left(\mathbf{w}\right)\right)$$

Some remarks

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

$$\begin{aligned}\min_{\Delta\mathbf{w}} \quad & \frac{1}{2}\Delta\mathbf{w}^\top H(\mathbf{w},\boldsymbol{\lambda},\boldsymbol{\mu})\Delta\mathbf{w} + \nabla\Phi\left(\mathbf{w}\right)^\top \Delta\mathbf{w} \\ \text{s.t.} \quad & \mathbf{g}\left(\mathbf{w}\right) + \nabla\mathbf{g}\left(\mathbf{w}\right)^\top \Delta\mathbf{w} = 0 \\ & \mathbf{h}\left(\mathbf{w}\right) + \nabla\mathbf{h}\left(\mathbf{w}\right)^\top \Delta\mathbf{w} \leq 0\end{aligned}$$

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

Some remarks

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

$$\begin{aligned} \min_{\Delta \mathbf{w}} \quad & \frac{1}{2}\Delta \mathbf{w}^\top H(\mathbf{w},\boldsymbol{\lambda},\boldsymbol{\mu})\Delta \mathbf{w} + \nabla \Phi\left(\mathbf{w}\right)^\top \Delta \mathbf{w} \\ \text{s.t.} \quad & \mathbf{g}\left(\mathbf{w}\right) + \nabla \mathbf{g}\left(\mathbf{w}\right)^\top \Delta \mathbf{w} = 0 \\ & \mathbf{h}\left(\mathbf{w}\right) + \nabla \mathbf{h}\left(\mathbf{w}\right)^\top \Delta \mathbf{w} \leq 0 \end{aligned}$$

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

- SQP inherits the convergence properties of the Newton method

Some remarks

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

$$\begin{aligned}\min_{\Delta\mathbf{w}} \quad & \frac{1}{2}\Delta\mathbf{w}^\top H(\mathbf{w},\boldsymbol{\lambda},\boldsymbol{\mu})\Delta\mathbf{w} + \nabla\Phi(\mathbf{w})^\top \Delta\mathbf{w} \\ \text{s.t.} \quad & \mathbf{g}(\mathbf{w}) + \nabla\mathbf{g}(\mathbf{w})^\top \Delta\mathbf{w} = 0 \\ & \mathbf{h}(\mathbf{w}) + \nabla\mathbf{h}(\mathbf{w})^\top \Delta\mathbf{w} \leq 0\end{aligned}$$

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

- SQP inherits the convergence properties of the Newton method
- What happens if SOSC fails during the iterations ? I.e. for an iterate $\mathbf{w},\boldsymbol{\lambda},\boldsymbol{\mu}$:

$$\mathbf{d}^\top H(\mathbf{w},\boldsymbol{\lambda},\boldsymbol{\mu})\mathbf{d} \not> 0$$

for some $\mathbf{d} \neq 0$ being a critical feasible direction ?

Some remarks

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

$$\begin{aligned} \min_{\Delta\mathbf{w}} \quad & \frac{1}{2}\Delta\mathbf{w}^\top H(\mathbf{w},\boldsymbol{\lambda},\boldsymbol{\mu})\Delta\mathbf{w} + \nabla\Phi\left(\mathbf{w}\right)^\top\Delta\mathbf{w} \\ \text{s.t.} \quad & \mathbf{g}\left(\mathbf{w}\right) + \nabla\mathbf{g}\left(\mathbf{w}\right)^\top\Delta\mathbf{w} = 0 \\ & \mathbf{h}\left(\mathbf{w}\right) + \nabla\mathbf{h}\left(\mathbf{w}\right)^\top\Delta\mathbf{w} \leq 0 \end{aligned}$$

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

- SQP inherits the convergence properties of the Newton method
- What happens if SOSC fails during the iterations ? I.e. for an iterate $\mathbf{w},\boldsymbol{\lambda},\boldsymbol{\mu}$:

  $$\mathbf{d}^\top H(\mathbf{w},\boldsymbol{\lambda},\boldsymbol{\mu})\mathbf{d} \not> 0$$

  for some $\mathbf{d} \neq 0$ being a critical feasible direction ? QP unbounded !! Heuristics are used in SQP methods to modify $H(\mathbf{w},\boldsymbol{\lambda},\boldsymbol{\mu})$ and recover an adequate curvature in the QP cost (regularization).

Outline

Newton-type Methods - Gauss-Newton Hessian approximation

$$\boxed{\text{Cost function of the type } \Phi(\mathbf{w}) = \frac{1}{2}\|\mathbf{R}(\mathbf{w})\|^2, \text{ with } \mathbf{R}(\mathbf{w}) \in \mathbb{R}^m}$$

Gauss-Newton Hessian approximation

Observe that

$$\nabla^2_{\mathbf{w}}\Phi(\mathbf{w}) = \frac{\partial}{\partial \mathbf{w}}\left(\nabla\mathbf{R}(\mathbf{w})\mathbf{R}(\mathbf{w})\right) = \nabla\mathbf{R}(\mathbf{w})\nabla\mathbf{R}(\mathbf{w})^\top + \sum_{i=1}^{m}\nabla^2\mathbf{R}_i(\mathbf{w})\mathbf{R}_i(\mathbf{w})$$

Newton-type Methods - Gauss-Newton Hessian approximation

$$\boxed{\text{Cost function of the type } \Phi(\mathbf{w}) = \frac{1}{2}\|\mathbf{R}(\mathbf{w})\|^2, \text{ with } \mathbf{R}(\mathbf{w}) \in \mathbb{R}^m}$$

Gauss-Newton Hessian approximation

Observe that

$$\nabla^2_{\mathbf{w}}\Phi(\mathbf{w}) = \frac{\partial}{\partial \mathbf{w}}\left(\nabla\mathbf{R}(\mathbf{w})\mathbf{R}(\mathbf{w})\right) = \nabla\mathbf{R}(\mathbf{w})\nabla\mathbf{R}(\mathbf{w})^\top + \sum_{i=1}^{m}\nabla^2\mathbf{R}_i(\mathbf{w})\mathbf{R}_i(\mathbf{w})$$

Gauss-Newton method proposes to use: $\boxed{B_k = \nabla\mathbf{R}(\mathbf{w}_k)\nabla\mathbf{R}(\mathbf{w}_k)^\top + \alpha_k \boldsymbol{I}}$

Newton-type Methods - Gauss-Newton Hessian approximation

$$\boxed{\text{Cost function of the type } \Phi(\mathbf{w}) = \frac{1}{2}\|\mathbf{R}(\mathbf{w})\|^2, \text{ with } \mathbf{R}(\mathbf{w}) \in \mathbb{R}^m}$$

Gauss-Newton Hessian approximation

Observe that

$$\nabla^2_{\mathbf{w}}\Phi(\mathbf{w}) = \frac{\partial}{\partial \mathbf{w}}\left(\nabla \mathbf{R}(\mathbf{w})\mathbf{R}(\mathbf{w})\right) = \nabla \mathbf{R}(\mathbf{w})\nabla \mathbf{R}(\mathbf{w})^\top + \sum_{i=1}^{m} \nabla^2 \mathbf{R}_i(\mathbf{w})\mathbf{R}_i(\mathbf{w})$$

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:

Newton-type Methods - Gauss-Newton Hessian approximation

$$\boxed{\text{Cost function of the type } \Phi(\mathbf{w}) = \frac{1}{2}\|\mathbf{R}(\mathbf{w})\|^2, \text{ with } \mathbf{R}(\mathbf{w}) \in \mathbb{R}^m}$$

Gauss-Newton Hessian approximation

Observe that

$$\nabla^2_{\mathbf{w}}\Phi(\mathbf{w}) = \frac{\partial}{\partial \mathbf{w}}\left(\nabla\mathbf{R}(\mathbf{w})\mathbf{R}(\mathbf{w})\right) = \nabla\mathbf{R}(\mathbf{w})\nabla\mathbf{R}(\mathbf{w})^\top + \sum_{i=1}^{m}\nabla^2\mathbf{R}_i(\mathbf{w})\mathbf{R}_i(\mathbf{w})$$

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:

- constraints are close to linear <u>or</u>
- $\Phi(\mathbf{w}^\star) \approx 0$ (implies $\boldsymbol{\lambda}$, $\boldsymbol{\mu} \approx 0$)

Newton-type Methods - Gauss-Newton Hessian approximation

$$\boxed{\text{Cost function of the type } \Phi(\mathbf{w}) = \frac{1}{2}\|\mathbf{R}(\mathbf{w})\|^2, \text{ with } \mathbf{R}(\mathbf{w}) \in \mathbb{R}^m}$$

Gauss-Newton Hessian approximation

Observe that

$$\nabla^2_{\mathbf{w}}\Phi\left(\mathbf{w}\right) = \frac{\partial}{\partial \mathbf{w}}\left(\nabla \mathbf{R}(\mathbf{w})\mathbf{R}(\mathbf{w})\right) = \nabla \mathbf{R}(\mathbf{w})\nabla \mathbf{R}(\mathbf{w})^\top + \sum_{i=1}^{m} \nabla^2 \mathbf{R}_i(\mathbf{w})\mathbf{R}_i(\mathbf{w})$$

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:

- constraints are close to linear or
- $\Phi(\mathbf{w}^\star) \approx 0$ (implies $\boldsymbol{\lambda}$, $\boldsymbol{\mu} \approx 0$) **and**

Newton-type Methods - Gauss-Newton Hessian approximation

$$\boxed{\text{Cost function of the type } \Phi(\mathbf{w}) = \frac{1}{2}\|\mathbf{R}(\mathbf{w})\|^2, \text{ with } \mathbf{R}(\mathbf{w}) \in \mathbb{R}^m}$$

Gauss-Newton Hessian approximation

Observe that

$$\nabla^2_{\mathbf{w}}\Phi(\mathbf{w}) = \frac{\partial}{\partial \mathbf{w}}\left(\nabla\mathbf{R}(\mathbf{w})\mathbf{R}(\mathbf{w})\right) = \nabla\mathbf{R}(\mathbf{w})\nabla\mathbf{R}(\mathbf{w})^\top + \sum_{i=1}^{m}\nabla^2\mathbf{R}_i(\mathbf{w})\mathbf{R}_i(\mathbf{w})$$

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:

- constraints are close to linear <u>or</u>
- $\Phi(\mathbf{w}^\star) \approx 0$ (implies $\boldsymbol{\lambda}$, $\boldsymbol{\mu} \approx 0$)

and

- all $\nabla^2\mathbf{R}_i(\mathbf{w})$ are small ($\mathbf{R}$ close to linear), <u>or</u>
- all $\mathbf{R}_i(\mathbf{w})$ are small, i.e. $\Phi(\mathbf{w}^\star) \approx 0$

Newton-type Methods - Gauss-Newton Hessian approximation

$$\boxed{\text{Cost function of the type } \Phi(\mathbf{w}) = \frac{1}{2}\|\mathbf{R}(\mathbf{w})\|^2, \text{ with } \mathbf{R}(\mathbf{w}) \in \mathbb{R}^m}$$

Gauss-Newton Hessian approximation

Observe that

$$\nabla^2_{\mathbf{w}}\Phi(\mathbf{w}) = \frac{\partial}{\partial \mathbf{w}}\left(\nabla\mathbf{R}(\mathbf{w})\mathbf{R}(\mathbf{w})\right) = \nabla\mathbf{R}(\mathbf{w})\nabla\mathbf{R}(\mathbf{w})^\top + \sum_{i=1}^{m}\nabla^2\mathbf{R}_i(\mathbf{w})\mathbf{R}_i(\mathbf{w})$$

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:

- constraints are close to linear <u>or</u>
- $\Phi(\mathbf{w}^\star) \approx 0$ (implies $\boldsymbol{\lambda}$, $\boldsymbol{\mu} \approx 0$)

and

- all $\nabla^2\mathbf{R}_i(\mathbf{w})$ are small ($\mathbf{R}$ close to linear), <u>or</u>
- all $\mathbf{R}_i(\mathbf{w})$ are small, i.e. $\Phi(\mathbf{w}^\star) \approx 0$

Typical application to tracking & fitting problems: $\mathbf{R}(\mathbf{w}) = \mathbf{y}(\mathbf{w}) - \bar{\mathbf{y}}$

Newton-type Methods - Gauss-Newton Hessian approximation

$$\boxed{\text{Cost function of the type } \Phi(\mathbf{w}) = \tfrac{1}{2}\|\mathbf{R}(\mathbf{w})\|^2, \text{ with } \mathbf{R}(\mathbf{w}) \in \mathbb{R}^m}$$

Gauss-Newton Hessian approximation

Observe that

$$\nabla^2_{\mathbf{w}}\Phi\left(\mathbf{w}\right) = \frac{\partial}{\partial \mathbf{w}}\left(\nabla\mathbf{R}(\mathbf{w})\mathbf{R}(\mathbf{w})\right) = \nabla\mathbf{R}(\mathbf{w})\nabla\mathbf{R}(\mathbf{w})^\top + \sum_{i=1}^{m}\nabla^2\mathbf{R}_i(\mathbf{w})\mathbf{R}_i(\mathbf{w})$$

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:

- constraints are close to linear <u>or</u>
- $\Phi(\mathbf{w}^\star) \approx 0$ (implies $\boldsymbol{\lambda}$, $\boldsymbol{\mu} \approx 0$)

and

- all $\nabla^2\mathbf{R}_i(\mathbf{w})$ are small ($\mathbf{R}$ close to linear), <u>or</u>
- all $\mathbf{R}_i(\mathbf{w})$ are small, i.e. $\Phi(\mathbf{w}^\star) \approx 0$

Typical application to tracking & fitting problems: $\mathbf{R}(\mathbf{w}) = \mathbf{y}(\mathbf{w}) - \bar{\mathbf{y}}$

Convergence

- If $\Phi\left(\mathbf{w}_k\right) \to 0$ then $\kappa_k \to 0$
- Can get superlinear convergence...

Newton-type Methods - BFGS

$\boxed{\text{Compute numerical derivative of } H(\mathbf{w}) \text{ in an efficient (iterative) way}}$

BFGS

Define

$$\begin{aligned}\mathbf{s}_k &= \mathbf{w}_{k+1} - \mathbf{w}_k \\ \mathbf{y}_k &= \nabla\mathcal{L}(\mathbf{w}_{k+1}) - \nabla\mathcal{L}(\mathbf{w}_k)\end{aligned}$$

Idea: $\boxed{\text{Update } B_k \rightarrow B_{k+1} \text{ such that } B_{k+1}\mathbf{s}_k = \mathbf{y}_k \text{ (secant condition)}}$

Newton-type Methods - BFGS

Compute numerical derivative of $H(\mathbf{w})$ in an efficient (iterative) way

BFGS

Define

$$\begin{aligned} \mathbf{s}_k &= \mathbf{w}_{k+1} - \mathbf{w}_k \\ \mathbf{y}_k &= \nabla\mathcal{L}(\mathbf{w}_{k+1}) - \nabla\mathcal{L}(\mathbf{w}_k) \end{aligned}$$

Idea: Update $B_k \rightarrow B_{k+1}$ such that $B_{k+1}\mathbf{s}_k = \mathbf{y}_k$ (secant condition)

BFGS formula: $$B_{k+1} = B_k - \frac{B_k \mathbf{s}\mathbf{s}^\top B_k}{\mathbf{s}^\top B_k \mathbf{s}} + \frac{\mathbf{y}\mathbf{y}^\top}{\mathbf{s}^\top \mathbf{y}}, \quad B_0 = I$$

Newton-type Methods - BFGS

$\boxed{\text{Compute numerical derivative of } H(\mathbf{w}) \text{ in an efficient (iterative) way}}$

BFGS

Define

$$\begin{aligned} \mathbf{s}_k &= \mathbf{w}_{k+1} - \mathbf{w}_k \\ \mathbf{y}_k &= \nabla\mathcal{L}(\mathbf{w}_{k+1}) - \nabla\mathcal{L}(\mathbf{w}_k) \end{aligned}$$

Idea: $\boxed{\text{Update } B_k \rightarrow B_{k+1} \text{ such that } B_{k+1}\mathbf{s}_k = \mathbf{y}_k \text{ (secant condition)}}$

BFGS formula: $$\boxed{B_{k+1} = B_k - \frac{B_k \mathbf{s}\mathbf{s}^\top B_k}{\mathbf{s}^\top B_k \mathbf{s}} + \frac{\mathbf{y}\mathbf{y}^\top}{\mathbf{s}^\top \mathbf{y}}, \quad B_0 = I}$$

See "Powell's trick" to make sure that $B_{k+1} > 0$

Newton-type Methods - BFGS

$\boxed{\text{Compute numerical derivative of } H(\mathbf{w}) \text{ in an efficient (iterative) way}}$

BFGS

Define

$$\begin{aligned} \mathbf{s}_k &= \mathbf{w}_{k+1} - \mathbf{w}_k \\ \mathbf{y}_k &= \nabla\mathcal{L}(\mathbf{w}_{k+1}) - \nabla\mathcal{L}(\mathbf{w}_k) \end{aligned}$$

Idea: $\boxed{\text{Update } B_k \rightarrow B_{k+1} \text{ such that } B_{k+1}\mathbf{s}_k = \mathbf{y}_k \text{ (secant condition)}}$

BFGS formula: $$\boxed{B_{k+1} = B_k - \frac{B_k \mathbf{s}\mathbf{s}^\top B_k}{\mathbf{s}^\top B_k \mathbf{s}} + \frac{\mathbf{y}\mathbf{y}^\top}{\mathbf{s}^\top \mathbf{y}}, \quad B_0 = I}$$

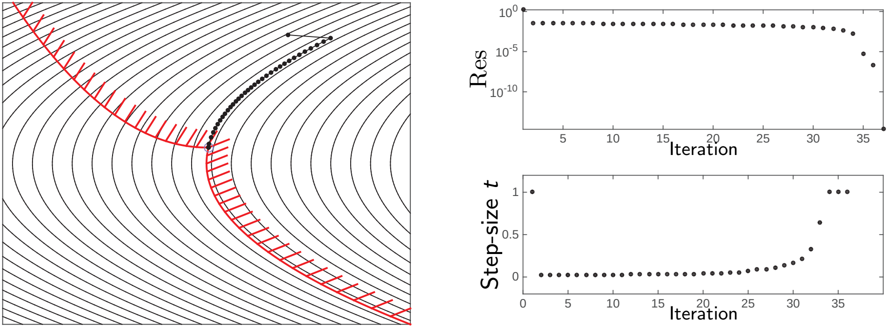
See "Powell's trick" to make sure that $B_{k+1} > 0$

Convergence

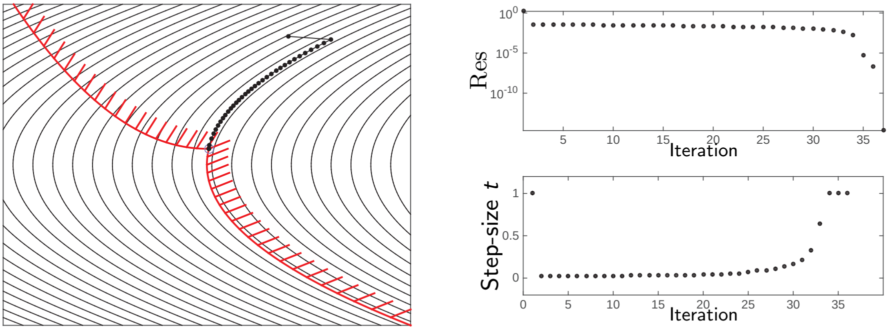
- It can be shown that $B_k \rightarrow H(\mathbf{w})$, then $\kappa_k \rightarrow 0$
- Can get superlinear convergence...

Outline

Maratos effect - Some NLPs can yield "creeping" convergence



Maratos effect - Some NLPs can yield "creeping" convergence



What is going on ?!? This is a case of the Maratos effect, can happen with nonlinear constraints...

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:

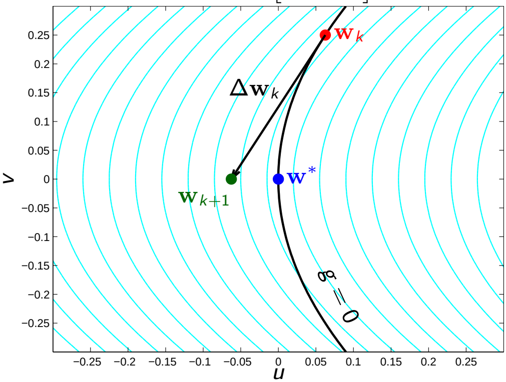
$$\mathbf{w}_k = \begin{bmatrix} a^2 & a \end{bmatrix}$$

The Newton step is:

$$\Delta\mathbf{w}_k = -\begin{bmatrix} 2a^2 & a \end{bmatrix}$$

for $\lambda = 2$...

Define: $\mathbf{w} = \begin{bmatrix} u & v \end{bmatrix}$



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:

$$\Delta\mathbf{w}_k = -\begin{bmatrix} 2a^2 & a \end{bmatrix}$$

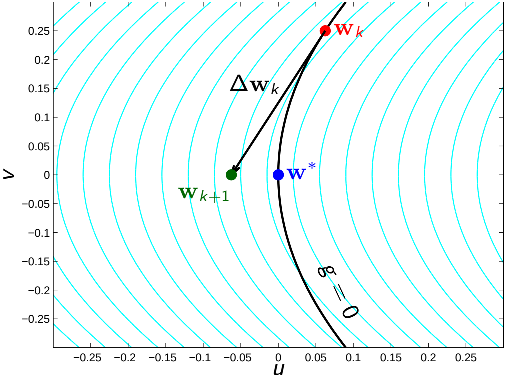
for $\lambda = 2$...

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: $\mathbf{w} = \begin{bmatrix} u & v \end{bmatrix}$



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:

$$\mathbf{w}_k = \begin{bmatrix} a^2 & a \end{bmatrix}$$

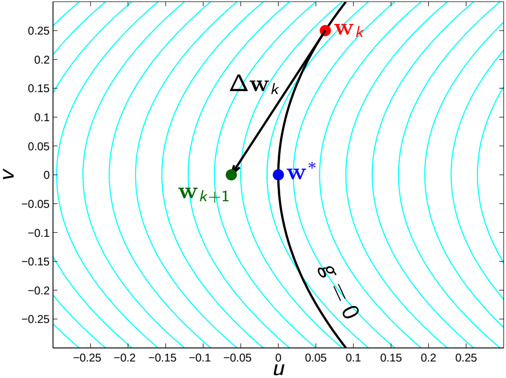
The Newton step is:

$$\Delta\mathbf{w}_k = -\begin{bmatrix} 2a^2 & a \end{bmatrix}$$

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 $\Delta\mathbf{w}_k$!!