

10 Karmarkars Algorithm

- ▶ inequalities $Ax \leq b$; $m \times n$ matrix A with rows a_i^T
- ▶ $P = \{x \mid Ax \leq b\}$; $P^\circ := \{x \mid Ax < b\}$
- ▶ interior point algorithm: $x \in P^\circ$ throughout the algorithm
- ▶ for $x \in P^\circ$ define

$$s_i(x) := b_i - a_i^T x$$

as the **slack** of the i -th constraint

logarithmic barrier function:

$$\phi(x) = - \sum_{i=1}^m \ln(s_i(x))$$

Penalty for point x ; points close to the boundary have a very large penalty.

Throughout this section a_i denotes the i -th row as a column vector.

10 Karmarkars Algorithm

- ▶ inequalities $Ax \leq b$; $m \times n$ matrix A with rows a_i^T
- ▶ $P = \{x \mid Ax \leq b\}$; $P^\circ := \{x \mid Ax < b\}$
- ▶ interior point algorithm: $x \in P^\circ$ throughout the algorithm
- ▶ for $x \in P^\circ$ define

$$s_i(x) := b_i - a_i^T x$$

as the **slack** of the i -th constraint

logarithmic barrier function:

$$\phi(x) = - \sum_{i=1}^m \ln(s_i(x))$$

Penalty for point x ; points close to the boundary have a very large penalty.

Throughout this section a_i denotes the i -th row as a column vector.

10 Karmarkars Algorithm

- ▶ inequalities $Ax \leq b$; $m \times n$ matrix A with rows a_i^T
- ▶ $P = \{x \mid Ax \leq b\}$; $P^\circ := \{x \mid Ax < b\}$
- ▶ interior point algorithm: $x \in P^\circ$ throughout the algorithm
- ▶ for $x \in P^\circ$ define

$$s_i(x) := b_i - a_i^T x$$

as the **slack** of the i -th constraint

logarithmic barrier function:

$$\phi(x) = - \sum_{i=1}^m \ln(s_i(x))$$

Penalty for point x ; points close to the boundary have a very large penalty.

Throughout this section a_i denotes the i -th row as a column vector.

10 Karmarkars Algorithm

- ▶ inequalities $Ax \leq b$; $m \times n$ matrix A with rows a_i^T
- ▶ $P = \{x \mid Ax \leq b\}$; $P^\circ := \{x \mid Ax < b\}$
- ▶ interior point algorithm: $x \in P^\circ$ throughout the algorithm
- ▶ for $x \in P^\circ$ define

$$s_i(x) := b_i - a_i^T x$$

as the **slack** of the i -th constraint

logarithmic barrier function:

$$\phi(x) = - \sum_{i=1}^m \ln(s_i(x))$$

Penalty for point x ; points close to the boundary have a very large penalty.

Throughout this section a_i denotes the i -th row as a column vector.

10 Karmarkars Algorithm

- ▶ inequalities $Ax \leq b$; $m \times n$ matrix A with rows a_i^T
- ▶ $P = \{x \mid Ax \leq b\}$; $P^\circ := \{x \mid Ax < b\}$
- ▶ interior point algorithm: $x \in P^\circ$ throughout the algorithm
- ▶ for $x \in P^\circ$ define

$$s_i(x) := b_i - a_i^T x$$

as the **slack** of the i -th constraint

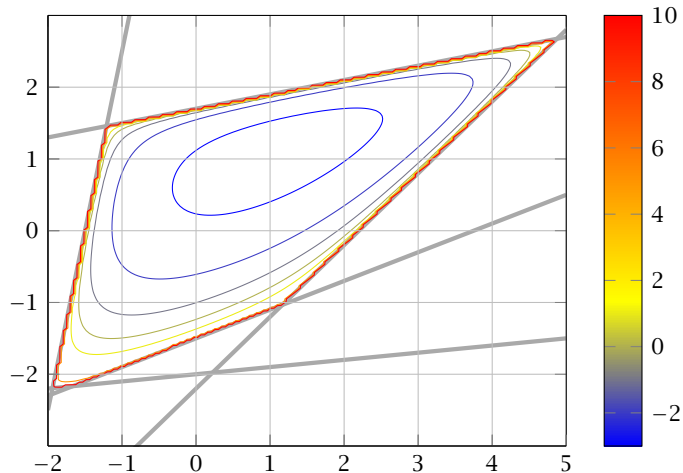
logarithmic barrier function:

$$\phi(x) = - \sum_{i=1}^m \ln(s_i(x))$$

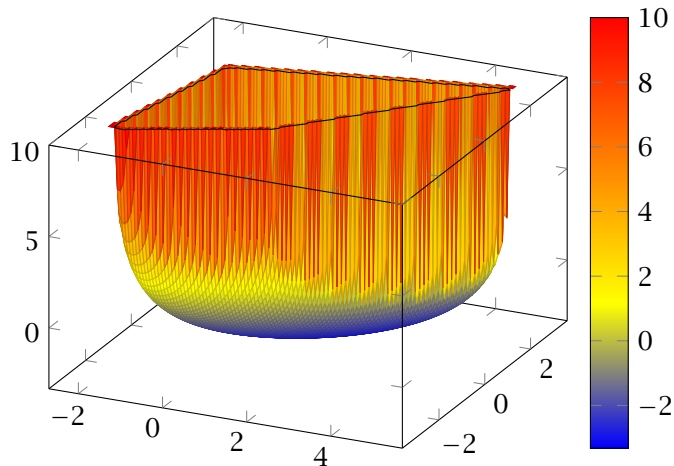
Penalty for point x ; points close to the boundary have a very large penalty.

Throughout this section a_i denotes the i -th row as a column vector.

Penalty Function



Penalty Function



Gradient and Hessian

Taylor approximation:

$$\phi(x + \epsilon) \approx \phi(x) + \nabla \phi(x)^T \epsilon + \frac{1}{2} \epsilon^T \nabla^2 \phi(x) \epsilon$$

Gradient:

$$\nabla \phi(x) = \sum_{i=1}^m \frac{1}{s_i(x)} \cdot a_i = A^T d_x$$

where $d_x^T = (1/s_1(x), \dots, 1/s_m(x))$. (d_x vector of inverse slacks)

Hessian:

$$H_x := \nabla^2 \phi(x) = \sum_{i=1}^m \frac{1}{s_i(x)^2} a_i a_i^T = A^T D_x^2 A$$

with $D_x = \text{diag}(d_x)$.

Gradient and Hessian

Taylor approximation:

$$\phi(x + \epsilon) \approx \phi(x) + \nabla \phi(x)^T \epsilon + \frac{1}{2} \epsilon^T \nabla^2 \phi(x) \epsilon$$

Gradient:

$$\nabla \phi(x) = \sum_{i=1}^m \frac{1}{s_i(x)} \cdot a_i = A^T d_x$$

where $d_x^T = (1/s_1(x), \dots, 1/s_m(x))$. (d_x vector of **inverse slacks**)

Hessian:

$$H_x := \nabla^2 \phi(x) = \sum_{i=1}^m \frac{1}{s_i(x)^2} a_i a_i^T = A^T D_x^2 A$$

with $D_x = \text{diag}(d_x)$.

Gradient and Hessian

Taylor approximation:

$$\phi(x + \epsilon) \approx \phi(x) + \nabla \phi(x)^T \epsilon + \frac{1}{2} \epsilon^T \nabla^2 \phi(x) \epsilon$$

Gradient:

$$\nabla \phi(x) = \sum_{i=1}^m \frac{1}{s_i(x)} \cdot a_i = A^T d_x$$

where $d_x^T = (1/s_1(x), \dots, 1/s_m(x))$. (d_x vector of **inverse slacks**)

Hessian:

$$H_x := \nabla^2 \phi(x) = \sum_{i=1}^m \frac{1}{s_i(x)^2} a_i a_i^T = A^T D_x^2 A$$

with $D_x = \text{diag}(d_x)$.

Proof for Gradient

$$\begin{aligned}\frac{\partial \phi(x)}{\partial x_i} &= \frac{\partial}{\partial x_i} \left(- \sum_r \ln(s_r(x)) \right) \\ &= - \sum_r \frac{\partial}{\partial x_i} \left(\ln(s_r(x)) \right) = - \sum_r \frac{1}{s_r(x)} \frac{\partial}{\partial x_i} \left(s_r(x) \right) \\ &= - \sum_r \frac{1}{s_r(x)} \frac{\partial}{\partial x_i} \left(b_r - a_r^T x \right) = \sum_r \frac{1}{s_r(x)} \frac{\partial}{\partial x_i} \left(a_r^T x \right) \\ &= \sum_r \frac{1}{s_r(x)} A_{ri}\end{aligned}$$

The i -th entry of the gradient vector is $\sum_r 1/s_r(x) \cdot A_{ri}$. This gives that the gradient is

$$\nabla \phi(x) = \sum_r \frac{1}{s_r(x)} a_r = A^T d_x$$

Proof for Hessian

$$\begin{aligned}\frac{\partial}{\partial x_j} \left(\sum_r \frac{1}{s_r(x)} A_{ri} \right) &= \sum_r A_{ri} \left(-\frac{1}{s_r(x)^2} \right) \cdot \frac{\partial}{\partial x_j} (s_r(x)) \\ &= \sum_r A_{ri} \frac{1}{s_r(x)^2} A_{rj}\end{aligned}$$

Note that $\sum_r A_{ri} A_{rj} = (A^T A)_{ij}$. Adding the additional factors $1/s_r(x)^2$ can be done with a diagonal matrix.

Hence the Hessian is

$$H_x = A^T D^2 A$$

Properties of the Hessian

H_x is positive semi-definite for $x \in P^\circ$

$$u^T H_x u = u^T A^T D_x^2 A u = \|D_x A u\|_2^2 \geq 0$$

This gives that $\phi(x)$ is convex.

If $\text{rank}(A) = n$, H_x is positive definite for $x \in P^\circ$

$$u^T H_x u = \|D_x A u\|_2^2 > 0 \text{ for } u \neq 0$$

This gives that $\phi(x)$ is **strictly** convex.

$\|u\|_{H_x} := \sqrt{u^T H_x u}$ is a (semi-)norm; the unit ball w.r.t. this norm is an ellipsoid.

Properties of the Hessian

H_x is positive semi-definite for $x \in P^\circ$

$$u^T H_x u = u^T A^T D_x^2 A u = \|D_x A u\|_2^2 \geq 0$$

This gives that $\phi(x)$ is convex.

If $\text{rank}(A) = n$, H_x is positive definite for $x \in P^\circ$

$$u^T H_x u = \|D_x A u\|_2^2 > 0 \text{ for } u \neq 0$$

This gives that $\phi(x)$ is **strictly** convex.

$\|u\|_{H_x} := \sqrt{u^T H_x u}$ is a (semi-)norm; the unit ball w.r.t. this norm is an ellipsoid.

Properties of the Hessian

H_x is positive semi-definite for $x \in P^\circ$

$$u^T H_x u = u^T A^T D_x^2 A u = \|D_x A u\|_2^2 \geq 0$$

This gives that $\phi(x)$ is convex.

If $\text{rank}(A) = n$, H_x is positive definite for $x \in P^\circ$

$$u^T H_x u = \|D_x A u\|_2^2 > 0 \text{ for } u \neq 0$$

This gives that $\phi(x)$ is strictly convex.

$\|u\|_{H_x} := \sqrt{u^T H_x u}$ is a (semi-)norm; the unit ball w.r.t. this norm is an ellipsoid.

Properties of the Hessian

H_x is positive semi-definite for $x \in P^\circ$

$$u^T H_x u = u^T A^T D_x^2 A u = \|D_x A u\|_2^2 \geq 0$$

This gives that $\phi(x)$ is convex.

If $\text{rank}(A) = n$, H_x is positive definite for $x \in P^\circ$

$$u^T H_x u = \|D_x A u\|_2^2 > 0 \text{ for } u \neq 0$$

This gives that $\phi(x)$ is **strictly** convex.

$\|u\|_{H_x} := \sqrt{u^T H_x u}$ is a (semi-)norm; the unit ball w.r.t. this norm is an ellipsoid.

Properties of the Hessian

H_x is positive semi-definite for $x \in P^\circ$

$$u^T H_x u = u^T A^T D_x^2 A u = \|D_x A u\|_2^2 \geq 0$$

This gives that $\phi(x)$ is convex.

If $\text{rank}(A) = n$, H_x is positive definite for $x \in P^\circ$

$$u^T H_x u = \|D_x A u\|_2^2 > 0 \text{ for } u \neq 0$$

This gives that $\phi(x)$ is **strictly** convex.

$\|u\|_{H_x} := \sqrt{u^T H_x u}$ is a (semi-)norm; the unit ball w.r.t. this norm is an ellipsoid.

Dikin Ellipsoid

$$E_x = \{y \mid (y - x)^T H_x (y - x) \leq 1\} = \{y \mid \|y - x\|_{H_x} \leq 1\}$$

Points in E_x are feasible!!!

Change of distance to 1st constraint going from x to y is $\frac{1}{2} \sum_{i=1}^m \frac{1}{\lambda_i} \frac{1}{\lambda_i} \frac{1}{\lambda_i} \frac{1}{\lambda_i} \dots$
Distance of x to 1st constraint

In order to become infeasible when going from x to y one of the terms in the sum would need to be larger than 1.

Dikin Ellipsoid

$$E_x = \{y \mid (y - x)^T H_x (y - x) \leq 1\} = \{y \mid \|y - x\|_{H_x} \leq 1\}$$

Points in E_x are feasible!!!

$$\begin{aligned} (y - x)^T H_x (y - x) &= (y - x)^T A^T D_x^2 A (y - x) \\ &= \sum_{i=1}^m \frac{(a_i^T (y - x))^2}{s_i(x)^2} \\ &= \sum_{i=1}^m \frac{(\text{change of distance to } i\text{-th constraint going from } x \text{ to } y)^2}{(\text{distance of } x \text{ to } i\text{-th constraint})^2} \\ &\leq 1 \end{aligned}$$

In order to become infeasible when going from x to y one of the terms in the sum would need to be larger than 1.

Dikin Ellipsoid

$$E_x = \{y \mid (y - x)^T H_x (y - x) \leq 1\} = \{y \mid \|y - x\|_{H_x} \leq 1\}$$

Points in E_x are feasible!!!

$$\begin{aligned} (y - x)^T H_x (y - x) &= (y - x)^T A^T D_x^2 A (y - x) \\ &= \sum_{i=1}^m \frac{(a_i^T (y - x))^2}{s_i(x)^2} \\ &= \sum_{i=1}^m \frac{(\text{change of distance to } i\text{-th constraint going from } x \text{ to } y)^2}{(\text{distance of } x \text{ to } i\text{-th constraint})^2} \\ &\leq 1 \end{aligned}$$

In order to become infeasible when going from x to y one of the terms in the sum would need to be larger than 1.

Dikin Ellipsoid

$$E_x = \{y \mid (y - x)^T H_x (y - x) \leq 1\} = \{y \mid \|y - x\|_{H_x} \leq 1\}$$

Points in E_x are feasible!!!

$$\begin{aligned} (y - x)^T H_x (y - x) &= (y - x)^T A^T D_x^2 A (y - x) \\ &= \sum_{i=1}^m \frac{(a_i^T (y - x))^2}{s_i(x)^2} \\ &= \sum_{i=1}^m \frac{(\text{change of distance to } i\text{-th constraint going from } x \text{ to } y)^2}{(\text{distance of } x \text{ to } i\text{-th constraint})^2} \\ &\leq 1 \end{aligned}$$

In order to become infeasible when going from x to y one of the terms in the sum would need to be larger than 1.

Dikin Ellipsoid

$$E_x = \{y \mid (y - x)^T H_x (y - x) \leq 1\} = \{y \mid \|y - x\|_{H_x} \leq 1\}$$

Points in E_x are feasible!!!

$$\begin{aligned}(y - x)^T H_x (y - x) &= (y - x)^T A^T D_x^2 A (y - x) \\ &= \sum_{i=1}^m \frac{(a_i^T (y - x))^2}{s_i(x)^2} \\ &= \sum_{i=1}^m \frac{(\text{change of distance to } i\text{-th constraint going from } x \text{ to } y)^2}{(\text{distance of } x \text{ to } i\text{-th constraint})^2} \\ &\leq 1\end{aligned}$$

In order to become infeasible when going from x to y one of the terms in the sum would need to be larger than 1.

Dikin Ellipsoid

$$E_x = \{y \mid (y - x)^T H_x (y - x) \leq 1\} = \{y \mid \|y - x\|_{H_x} \leq 1\}$$

Points in E_x are feasible!!!

$$\begin{aligned} (y - x)^T H_x (y - x) &= (y - x)^T A^T D_x^2 A (y - x) \\ &= \sum_{i=1}^m \frac{(a_i^T (y - x))^2}{s_i(x)^2} \\ &= \sum_{i=1}^m \frac{(\text{change of distance to } i\text{-th constraint going from } x \text{ to } y)^2}{(\text{distance of } x \text{ to } i\text{-th constraint})^2} \\ &\leq 1 \end{aligned}$$

In order to become infeasible when going from x to y one of the terms in the sum would need to be larger than 1.

Dikin Ellipsoid

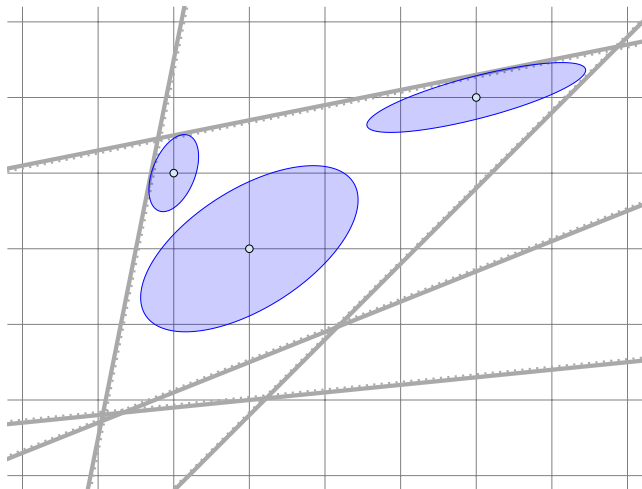
$$E_x = \{y \mid (y - x)^T H_x (y - x) \leq 1\} = \{y \mid \|y - x\|_{H_x} \leq 1\}$$

Points in E_x are feasible!!!

$$\begin{aligned} (y - x)^T H_x (y - x) &= (y - x)^T A^T D_x^2 A (y - x) \\ &= \sum_{i=1}^m \frac{(a_i^T (y - x))^2}{s_i(x)^2} \\ &= \sum_{i=1}^m \frac{(\text{change of distance to } i\text{-th constraint going from } x \text{ to } y)^2}{(\text{distance of } x \text{ to } i\text{-th constraint})^2} \\ &\leq 1 \end{aligned}$$

In order to become infeasible when going from x to y one of the terms in the sum would need to be larger than 1.

Dikin Ellipsoids



$$x_{ac} := \arg \min_{x \in P^\circ} \phi(x)$$

- ▶ x_{ac} is solution to

$$\nabla \phi(x) = \sum_{i=1}^m \frac{1}{s_i(x)} a_i = 0$$

- ▶ depends on the **description** of the polytope
- ▶ x_{ac} exists and is unique iff P° is nonempty and bounded

Central Path

In the following we assume that the LP and its dual are **strictly feasible** and that $\text{rank}(A) = n$.

Central Path:

Set of points $\{x^*(t) \mid t > 0\}$ with

$$x^*(t) = \operatorname{argmin}_x \{tc^T x + \phi(x)\}$$

- ▶ $t = 0$: analytic center
- ▶ $t = \infty$: optimum solution

$x^*(t)$ exists and is unique for all $t \geq 0$.

Central Path

In the following we assume that the LP and its dual are **strictly feasible** and that $\text{rank}(A) = n$.

Central Path:

Set of points $\{x^*(t) \mid t > 0\}$ with

$$x^*(t) = \operatorname{argmin}_x \{tc^T x + \phi(x)\}$$

- ▶ $t = 0$: analytic center
- ▶ $t = \infty$: optimum solution

$x^*(t)$ exists and is unique for all $t \geq 0$.

Central Path

In the following we assume that the LP and its dual are **strictly feasible** and that $\text{rank}(A) = n$.

Central Path:

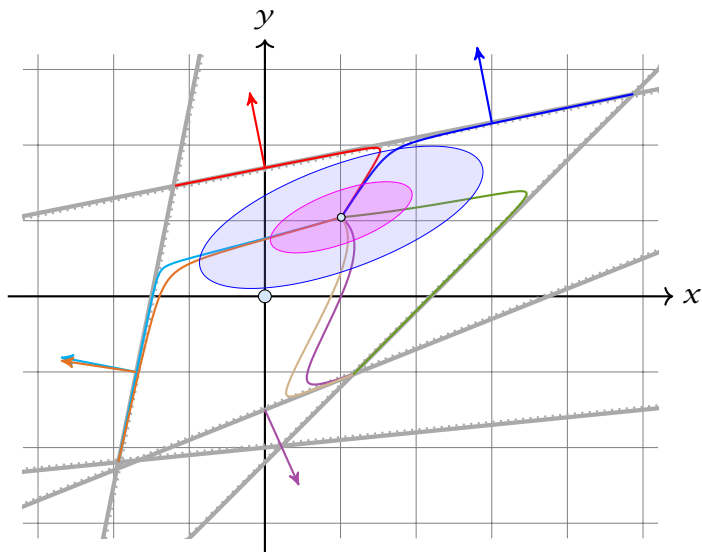
Set of points $\{x^*(t) \mid t > 0\}$ with

$$x^*(t) = \operatorname{argmin}_x \{tc^T x + \phi(x)\}$$

- ▶ $t = 0$: analytic center
- ▶ $t = \infty$: optimum solution

$x^*(t)$ exists and is unique for all $t \geq 0$.

Different Central Paths



Central Path

Intuitive Idea:

Find point on central path for large value of t . Should be close to optimum solution.

Questions:

- ▶ Is this really true? How large a t do we need?
- ▶ How do we find corresponding point $x^*(t)$ on central path?

The Dual

primal-dual pair:

$$\begin{array}{ll} \min & c^T x \\ \text{s.t.} & Ax \leq b \end{array}$$

$$\begin{array}{ll} \max & -b^T z \\ \text{s.t.} & A^T z + c = 0 \\ & z \geq 0 \end{array}$$

Assumptions

- ▶ primal and dual problems are strictly feasible;
- ▶ $\text{rank}(A) = n$.

Note that the right LP in standard form is equal to $\max\{-b^T y \mid -A^T y = c, x \geq 0\}$. The dual of this is $\min\{c^T x \mid -Ax \geq -b\}$ (variables x are unrestricted).

Force Field Interpretation

Point $x^*(t)$ on central path is solution to $tc + \nabla\phi(x) = 0$

- ▶ We can view each constraint as generating a repelling force. The combination of these forces is represented by $\nabla\phi(x)$.
- ▶ In addition there is a force tc pulling us towards the optimum solution.

How large should t be?

Point $x^*(t)$ on central path is solution to $tc + \nabla\phi(x) = 0$.

This means

$$tc + \sum_{i=1}^m \frac{1}{s_i(x^*(t))} a_i = 0$$

or

$$c + \sum_{i=1}^m z_i^*(t) a_i = 0 \quad \text{with} \quad z_i^*(t) = \frac{1}{ts_i(x^*(t))}$$

$(z^*(t), x^*(t))$ is strictly dual feasible, $x^*(t)$ is strictly primal feasible.

Quality gap between $(z^*(t), x^*(t))$ and (z^*, x^*) is

$\frac{1}{t}$ (primal gap) and $\frac{1}{t}$ (dual gap) respectively.

How large should t be?

Point $x^*(t)$ on central path is solution to $tc + \nabla\phi(x) = 0$.

This means

$$tc + \sum_{i=1}^m \frac{1}{s_i(x^*(t))} a_i = 0$$

or

$$c + \sum_{i=1}^m z_i^*(t) a_i = 0 \quad \text{with} \quad z_i^*(t) = \frac{1}{ts_i(x^*(t))}$$

Primal is strictly dual feasible.

Quality gap between primal and dual is $\frac{1}{t}$.

Primal is strictly dual feasible and the quality gap between primal and dual is $\frac{1}{t}$.

How large should t be?

Point $x^*(t)$ on central path is solution to $tc + \nabla\phi(x) = 0$.

This means

$$tc + \sum_{i=1}^m \frac{1}{s_i(x^*(t))} a_i = 0$$

or

$$c + \sum_{i=1}^m z_i^*(t) a_i = 0 \quad \text{with} \quad z_i^*(t) = \frac{1}{ts_i(x^*(t))}$$

How large should t be?

Point $x^*(t)$ on central path is solution to $tc + \nabla\phi(x) = 0$.

This means

$$tc + \sum_{i=1}^m \frac{1}{s_i(x^*(t))} a_i = 0$$

or

$$c + \sum_{i=1}^m z_i^*(t) a_i = 0 \quad \text{with} \quad z_i^*(t) = \frac{1}{ts_i(x^*(t))}$$

- ▶ $z^*(t)$ is strictly dual feasible: $(A^T z^* + c = 0; z^* > 0)$
- ▶ duality gap between $x := x^*(t)$ and $z := z^*(t)$ is

$$c^T x + b^T z = (b - Ax)^T z = \frac{m}{t}$$

- ▶ if gap is less than $1/2^{\Omega(L)}$ we can snap to optimum point

How large should t be?

Point $x^*(t)$ on central path is solution to $tc + \nabla\phi(x) = 0$.

This means

$$tc + \sum_{i=1}^m \frac{1}{s_i(x^*(t))} a_i = 0$$

or

$$c + \sum_{i=1}^m z_i^*(t) a_i = 0 \quad \text{with} \quad z_i^*(t) = \frac{1}{ts_i(x^*(t))}$$

- ▶ $z^*(t)$ is strictly dual feasible: $(A^T z^* + c = 0; z^* > 0)$
- ▶ duality gap between $x := x^*(t)$ and $z := z^*(t)$ is

$$c^T x + b^T z = (b - Ax)^T z = \frac{m}{t}$$

- ▶ if gap is less than $1/2^{\Omega(L)}$ we can snap to optimum point

How large should t be?

Point $x^*(t)$ on central path is solution to $tc + \nabla\phi(x) = 0$.

This means

$$tc + \sum_{i=1}^m \frac{1}{s_i(x^*(t))} a_i = 0$$

or

$$c + \sum_{i=1}^m z_i^*(t) a_i = 0 \quad \text{with} \quad z_i^*(t) = \frac{1}{ts_i(x^*(t))}$$

- ▶ $z^*(t)$ is strictly dual feasible: $(A^T z^* + c = 0; z^* > 0)$
- ▶ duality gap between $x := x^*(t)$ and $z := z^*(t)$ is

$$c^T x + b^T z = (b - Ax)^T z = \frac{m}{t}$$

- ▶ if gap is less than $1/2^{\Omega(L)}$ we can snap to optimum point

How to find $x^*(t)$

First idea:

- ▶ start somewhere in the polytope
- ▶ use iterative method (**Newtons method**) to minimize $f_t(x) := tc^T x + \phi(x)$

Newton Method

Quadratic approximation of f_t

$$f_t(\mathbf{x} + \epsilon) \approx f_t(\mathbf{x}) + \nabla f_t(\mathbf{x})^T \epsilon + \frac{1}{2} \epsilon^T H_{f_t}(\mathbf{x}) \epsilon$$

Suppose this were exact:

$$f_t(\mathbf{x} + \epsilon) = f_t(\mathbf{x}) + \nabla f_t(\mathbf{x})^T \epsilon + \frac{1}{2} \epsilon^T H_{f_t}(\mathbf{x}) \epsilon$$

Then gradient is given by:

$$\nabla f_t(\mathbf{x} + \epsilon) = \nabla f_t(\mathbf{x}) + H_{f_t}(\mathbf{x}) \cdot \epsilon$$

Note that for the one-dimensional case $g(\epsilon) = f(x) + f'(x)\epsilon + \frac{1}{2}f''(x)\epsilon^2$, then $g'(\epsilon) = f'(x) + f''(x)\epsilon$.

Newton Method

Quadratic approximation of f_t

$$f_t(\mathbf{x} + \epsilon) \approx f_t(\mathbf{x}) + \nabla f_t(\mathbf{x})^T \epsilon + \frac{1}{2} \epsilon^T H_{f_t}(\mathbf{x}) \epsilon$$

Suppose this were exact:

$$f_t(\mathbf{x} + \epsilon) = f_t(\mathbf{x}) + \nabla f_t(\mathbf{x})^T \epsilon + \frac{1}{2} \epsilon^T H_{f_t}(\mathbf{x}) \epsilon$$

Then gradient is given by:

$$\nabla f_t(\mathbf{x} + \epsilon) = \nabla f_t(\mathbf{x}) + H_{f_t}(\mathbf{x}) \cdot \epsilon$$

Note that for the one-dimensional case $g(\epsilon) = f(x) + f'(x)\epsilon + \frac{1}{2}f''(x)\epsilon^2$, then $g'(\epsilon) = f'(x) + f''(x)\epsilon$.

Newton Method

Quadratic approximation of f_t

$$f_t(\mathbf{x} + \epsilon) \approx f_t(\mathbf{x}) + \nabla f_t(\mathbf{x})^T \epsilon + \frac{1}{2} \epsilon^T H_{f_t}(\mathbf{x}) \epsilon$$

Suppose this were exact:

$$f_t(\mathbf{x} + \epsilon) = f_t(\mathbf{x}) + \nabla f_t(\mathbf{x})^T \epsilon + \frac{1}{2} \epsilon^T H_{f_t}(\mathbf{x}) \epsilon$$

Then gradient is given by:

$$\nabla f_t(\mathbf{x} + \epsilon) = \nabla f_t(\mathbf{x}) + H_{f_t}(\mathbf{x}) \cdot \epsilon$$

Note that for the one-dimensional case $g(\epsilon) = f(x) + f'(x)\epsilon + \frac{1}{2}f''(x)\epsilon^2$, then $g'(\epsilon) = f'(x) + f''(x)\epsilon$.

Newton Method

Observe that $H_{f_t}(x) = H(x)$, where $H(x)$ is the Hessian for the function $\phi(x)$ (adding a linear term like $tc^T x$ does not affect the Hessian).

Also $\nabla f_t(x) = tc + \nabla \phi(x)$.

We want to move to a point where this gradient is 0 :

Newton Step at $x \in P^\circ$

$$\begin{aligned}\Delta x_{nt} &= -H_{f_t}^{-1}(x) \nabla f_t(x) \\ &= -H_{f_t}^{-1}(x) (tc + \nabla \phi(x)) \\ &= -(A^T D_x^2 A)^{-1} (tc + A^T d_x)\end{aligned}$$

Newton Iteration:

$$x := x + \Delta x_{nt}$$

Measuring Progress of Newton Step

Newton decrement:

$$\begin{aligned}\lambda_t(x) &= \|D_x A \Delta x_{nt}\| \\ &= \|\Delta x_{nt}\|_{H_x}\end{aligned}$$

Square of Newton decrement is linear estimate of reduction if we do a Newton step:

$$-\lambda_t(x)^2 = \nabla f_t(x)^T \Delta x_{nt}$$

- ▶ $\lambda_t(x) = 0$ iff $x = x^*(t)$
- ▶ $\lambda_t(x)$ is measure of proximity of x to $x^*(t)$

Recall that Δx_{nt} fulfills $-H(x)\Delta x_{nt} = \nabla f_t(\cdot)$.

Measuring Progress of Newton Step

Newton decrement:

$$\begin{aligned}\lambda_t(\mathbf{x}) &= \|D_x A \Delta \mathbf{x}_{nt}\| \\ &= \|\Delta \mathbf{x}_{nt}\|_{H_x}\end{aligned}$$

Square of Newton decrement is linear estimate of reduction if we do a Newton step:

$$-\lambda_t(\mathbf{x})^2 = \nabla f_t(\mathbf{x})^T \Delta \mathbf{x}_{nt}$$

- ▶ $\lambda_t(\mathbf{x}) = 0$ iff $\mathbf{x} = \mathbf{x}^*(t)$
- ▶ $\lambda_t(\mathbf{x})$ is measure of proximity of \mathbf{x} to $\mathbf{x}^*(t)$

Recall that $\Delta \mathbf{x}_{nt}$ fulfills $-H(\mathbf{x})\Delta \mathbf{x}_{nt} = \nabla f_t(\cdot)$.

Measuring Progress of Newton Step

Newton decrement:

$$\begin{aligned}\lambda_t(\mathbf{x}) &= \|D_x A \Delta \mathbf{x}_{nt}\| \\ &= \|\Delta \mathbf{x}_{nt}\|_{H_x}\end{aligned}$$

Square of Newton decrement is linear estimate of reduction if we do a Newton step:

$$-\lambda_t(\mathbf{x})^2 = \nabla f_t(\mathbf{x})^T \Delta \mathbf{x}_{nt}$$

- ▶ $\lambda_t(\mathbf{x}) = 0$ iff $\mathbf{x} = \mathbf{x}^*(t)$
- ▶ $\lambda_t(\mathbf{x})$ is measure of proximity of \mathbf{x} to $\mathbf{x}^*(t)$

Recall that $\Delta \mathbf{x}_{nt}$ fulfills $-H(\mathbf{x})\Delta \mathbf{x}_{nt} = \nabla f_t(\cdot)$.

Convergence of Newtons Method

Theorem 2

If $\lambda_t(x) < 1$ then

- ▶ $x_+ := x + \Delta x_{nt} \in P^\circ$ (new point feasible)
- ▶ $\lambda_t(x_+) \leq \lambda_t(x)^2$

This means we have **quadratic convergence**. Very fast.

Convergence of Newtons Method

feasibility:

- ▶ $\lambda_t(\mathbf{x}) = \|\Delta\mathbf{x}_{nt}\|_{H_x} < 1$; hence \mathbf{x}_+ lies in the **Dikin ellipsoid** around \mathbf{x} .

Convergence of Newtons Method

bound on $\lambda_t(\mathbf{x}^+)$:

we use $D := D_x = \text{diag}(d_x)$ and $D_+ := D_{x^+} = \text{diag}(d_{x^+})$

To see the last equality we use Pythagoras

$$\|a\|^2 + \|a + b\|^2 = \|b\|^2$$

if $a^T(a + b) = 0$.

Convergence of Newtons Method

bound on $\lambda_t(\mathbf{x}^+)$:

we use $D := D_x = \text{diag}(d_x)$ and $D_+ := D_{x^+} = \text{diag}(d_{x^+})$

$$\begin{aligned}\lambda_t(\mathbf{x}^+)^2 &= \|D_+ A \Delta x_{\text{nt}}^+\|^2 \\ &\leq \|D_+ A \Delta x_{\text{nt}}^+\|^2 + \|D_+ A \Delta x_{\text{nt}}^+ + (I - D_+^{-1} D) D A \Delta x_{\text{nt}}\|^2 \\ &= \|(I - D_+^{-1} D) D A \Delta x_{\text{nt}}\|^2\end{aligned}$$

To see the last equality we use Pythagoras

$$\|a\|^2 + \|a + b\|^2 = \|b\|^2$$

if $a^T(a + b) = 0$.

Convergence of Newtons Method

bound on $\lambda_t(\mathbf{x}^+)$:

we use $D := D_x = \text{diag}(d_x)$ and $D_+ := D_{x^+} = \text{diag}(d_{x^+})$

$$\begin{aligned}\lambda_t(\mathbf{x}^+)^2 &= \|D_+ A \Delta \mathbf{x}_{\text{nt}}^+\|^2 \\ &\leq \|D_+ A \Delta \mathbf{x}_{\text{nt}}^+\|^2 + \|D_+ A \Delta \mathbf{x}_{\text{nt}}^+ + (I - D_+^{-1} D) D A \Delta \mathbf{x}_{\text{nt}}\|^2 \\ &= \|(I - D_+^{-1} D) D A \Delta \mathbf{x}_{\text{nt}}\|^2\end{aligned}$$

To see the last equality we use Pythagoras

$$\|a\|^2 + \|a + b\|^2 = \|b\|^2$$

if $a^T(a + b) = 0$.

Convergence of Newtons Method

bound on $\lambda_t(\mathbf{x}^+)$:

we use $D := D_x = \text{diag}(d_x)$ and $D_+ := D_{x^+} = \text{diag}(d_{x^+})$

$$\begin{aligned}\lambda_t(\mathbf{x}^+)^2 &= \|D_+ A \Delta \mathbf{x}_{\text{nt}}^+\|^2 \\ &\leq \|D_+ A \Delta \mathbf{x}_{\text{nt}}^+\|^2 + \|D_+ A \Delta \mathbf{x}_{\text{nt}}^+ + (I - D_+^{-1} D) D A \Delta \mathbf{x}_{\text{nt}}\|^2 \\ &= \|(I - D_+^{-1} D) D A \Delta \mathbf{x}_{\text{nt}}\|^2\end{aligned}$$

To see the last equality we use Pythagoras

$$\|a\|^2 + \|a + b\|^2 = \|b\|^2$$

if $a^T(a + b) = 0$.

Convergence of Newtons Method

bound on $\lambda_t(\mathbf{x}^+)$:

we use $D := D_x = \text{diag}(d_x)$ and $D_+ := D_{x^+} = \text{diag}(d_{x^+})$

$$\begin{aligned}\lambda_t(\mathbf{x}^+)^2 &= \|D_+ A \Delta \mathbf{x}_{\text{nt}}^+\|^2 \\ &\leq \|D_+ A \Delta \mathbf{x}_{\text{nt}}^+\|^2 + \|D_+ A \Delta \mathbf{x}_{\text{nt}}^+ + (I - D_+^{-1} D) D A \Delta \mathbf{x}_{\text{nt}}\|^2 \\ &= \|(I - D_+^{-1} D) D A \Delta \mathbf{x}_{\text{nt}}\|^2\end{aligned}$$

To see the last equality we use Pythagoras

$$\|a\|^2 + \|a + b\|^2 = \|b\|^2$$

if $a^T(a + b) = 0$.

Convergence of Newtons Method

bound on $\lambda_t(\mathbf{x}^+)$:

we use $D := D_x = \text{diag}(d_x)$ and $D_+ := D_{x^+} = \text{diag}(d_{x^+})$

$$\begin{aligned}\lambda_t(\mathbf{x}^+)^2 &= \|D_+ A \Delta \mathbf{x}_{\text{nt}}^+\|^2 \\ &\leq \|D_+ A \Delta \mathbf{x}_{\text{nt}}^+\|^2 + \|D_+ A \Delta \mathbf{x}_{\text{nt}}^+ + (I - D_+^{-1} D) D A \Delta \mathbf{x}_{\text{nt}}\|^2 \\ &= \|(I - D_+^{-1} D) D A \Delta \mathbf{x}_{\text{nt}}\|^2\end{aligned}$$

To see the last equality we use Pythagoras

$$\|a\|^2 + \|a + b\|^2 = \|b\|^2$$

if $a^T(a + b) = 0$.

Convergence of Newtons Method

$$\begin{aligned}DA\Delta x_{nt} &= DA(x^+ - x) \\ &= D(b - Ax - (b - Ax^+)) \\ &= D(D^{-1}\vec{1} - D_+^{-1}\vec{1}) \\ &= (I - D_+^{-1}D)\vec{1}\end{aligned}$$

$$\begin{aligned}a^T(a + b) &= \Delta x_{nt}^{+T} A^T D_+ (D_+ A \Delta x_{nt}^+ + (I - D_+^{-1} D) DA \Delta x_{nt}) \\ &= \Delta x_{nt}^{+T} (A^T D_+^2 A \Delta x_{nt}^+ - A^T D^2 A \Delta x_{nt} + A^T D_+ DA \Delta x_{nt}) \\ &= \Delta x_{nt}^{+T} (H_+ \Delta x_{nt}^+ - H \Delta x_{nt} + A^T D_+ \vec{1} - A^T D \vec{1}) \\ &= \Delta x_{nt}^{+T} (-\nabla f_t(x^+) + \nabla f_t(x) + \nabla \phi(x^+) - \nabla \phi(x)) \\ &= 0\end{aligned}$$

Convergence of Newtons Method

$$\begin{aligned}DA\Delta x_{nt} &= DA(x^+ - x) \\ &= D(b - Ax - (b - Ax^+)) \\ &= D(D^{-1}\vec{1} - D_+^{-1}\vec{1}) \\ &= (I - D_+^{-1}D)\vec{1}\end{aligned}$$

$$\begin{aligned}a^T(a + b) &= \Delta x_{nt}^{+T} A^T D_+ (D_+ A \Delta x_{nt}^+ + (I - D_+^{-1} D) DA \Delta x_{nt}) \\ &= \Delta x_{nt}^{+T} (A^T D_+^2 A \Delta x_{nt}^+ - A^T D^2 A \Delta x_{nt} + A^T D_+ DA \Delta x_{nt}) \\ &= \Delta x_{nt}^{+T} (H_+ \Delta x_{nt}^+ - H \Delta x_{nt} + A^T D_+ \vec{1} - A^T D \vec{1}) \\ &= \Delta x_{nt}^{+T} (-\nabla f_t(x^+) + \nabla f_t(x) + \nabla \phi(x^+) - \nabla \phi(x)) \\ &= 0\end{aligned}$$

Convergence of Newtons Method

$$\begin{aligned}DA\Delta x_{nt} &= DA(x^+ - x) \\ &= D(b - Ax - (b - Ax^+)) \\ &= D(D^{-1}\bar{1} - D_+^{-1}\bar{1}) \\ &= (I - D_+^{-1}D)\bar{1}\end{aligned}$$

$$\begin{aligned}a^T(a + b) &= \Delta x_{nt}^{+T} A^T D_+ \left(D_+ A \Delta x_{nt}^+ + (I - D_+^{-1}D) DA \Delta x_{nt} \right) \\ &= \Delta x_{nt}^{+T} \left(A^T D_+^2 A \Delta x_{nt}^+ - A^T D^2 A \Delta x_{nt} + A^T D_+ DA \Delta x_{nt} \right) \\ &= \Delta x_{nt}^{+T} \left(H_+ \Delta x_{nt}^+ - H \Delta x_{nt} + A^T D_+ \bar{1} - A^T D \bar{1} \right) \\ &= \Delta x_{nt}^{+T} \left(-\nabla f_t(x^+) + \nabla f_t(x) + \nabla \phi(x^+) - \nabla \phi(x) \right) \\ &= 0\end{aligned}$$

Convergence of Newtons Method

$$\begin{aligned}DA\Delta x_{nt} &= DA(x^+ - x) \\ &= D(b - Ax - (b - Ax^+)) \\ &= D(D^{-1}\vec{1} - D_+^{-1}\vec{1}) \\ &= (I - D_+^{-1}D)\vec{1}\end{aligned}$$

$$\begin{aligned}a^T(a + b) &= \Delta x_{nt}^{+T} A^T D_+ (D_+ A \Delta x_{nt}^+ + (I - D_+^{-1} D) DA \Delta x_{nt}) \\ &= \Delta x_{nt}^{+T} (A^T D_+^2 A \Delta x_{nt}^+ - A^T D^2 A \Delta x_{nt} + A^T D_+ DA \Delta x_{nt}) \\ &= \Delta x_{nt}^{+T} (H_+ \Delta x_{nt}^+ - H \Delta x_{nt} + A^T D_+ \vec{1} - A^T D \vec{1}) \\ &= \Delta x_{nt}^{+T} (-\nabla f_t(x^+) + \nabla f_t(x) + \nabla \phi(x^+) - \nabla \phi(x)) \\ &= 0\end{aligned}$$

Convergence of Newtons Method

$$\begin{aligned}DA\Delta x_{nt} &= DA(x^+ - x) \\ &= D(b - Ax - (b - Ax^+)) \\ &= D(D^{-1}\vec{1} - D_+^{-1}\vec{1}) \\ &= (I - D_+^{-1}D)\vec{1}\end{aligned}$$

$$\begin{aligned}a^T(a + b) &= \Delta x_{nt}^{+T} A^T D_+ (D_+ A \Delta x_{nt}^+ + (I - D_+^{-1} D) DA \Delta x_{nt}) \\ &= \Delta x_{nt}^{+T} (A^T D_+^2 A \Delta x_{nt}^+ - A^T D^2 A \Delta x_{nt} + A^T D_+ DA \Delta x_{nt}) \\ &= \Delta x_{nt}^{+T} (H_+ \Delta x_{nt}^+ - H \Delta x_{nt} + A^T D_+ \vec{1} - A^T D \vec{1}) \\ &= \Delta x_{nt}^{+T} (-\nabla f_t(x^+) + \nabla f_t(x) + \nabla \phi(x^+) - \nabla \phi(x)) \\ &= 0\end{aligned}$$

Convergence of Newtons Method

$$\begin{aligned}DA\Delta x_{nt} &= DA(x^+ - x) \\ &= D(b - Ax - (b - Ax^+)) \\ &= D(D^{-1}\vec{1} - D_+^{-1}\vec{1}) \\ &= (I - D_+^{-1}D)\vec{1}\end{aligned}$$

$$a^T(a + b)$$

$$\begin{aligned}&= \Delta x_{nt}^{+T} A^T D_+ (D_+ A \Delta x_{nt}^+ + (I - D_+^{-1} D) DA \Delta x_{nt}) \\ &= \Delta x_{nt}^{+T} (A^T D_+^2 A \Delta x_{nt}^+ - A^T D^2 A \Delta x_{nt} + A^T D_+ DA \Delta x_{nt}) \\ &= \Delta x_{nt}^{+T} (H_+ \Delta x_{nt}^+ - H \Delta x_{nt} + A^T D_+ \vec{1} - A^T D \vec{1}) \\ &= \Delta x_{nt}^{+T} (-\nabla f_t(x^+) + \nabla f_t(x) + \nabla \phi(x^+) - \nabla \phi(x)) \\ &= 0\end{aligned}$$

Convergence of Newtons Method

$$\begin{aligned}DA\Delta x_{nt} &= DA(x^+ - x) \\ &= D(b - Ax - (b - Ax^+)) \\ &= D(D^{-1}\vec{1} - D_+^{-1}\vec{1}) \\ &= (I - D_+^{-1}D)\vec{1}\end{aligned}$$

$$\begin{aligned}a^T(a + b) &= \Delta x_{nt}^{+T} A^T D_+ (D_+ A \Delta x_{nt}^+ + (I - D_+^{-1} D) D A \Delta x_{nt}) \\ &= \Delta x_{nt}^{+T} (A^T D_+^2 A \Delta x_{nt}^+ - A^T D^2 A \Delta x_{nt} + A^T D_+ D A \Delta x_{nt}) \\ &= \Delta x_{nt}^{+T} (H_+ \Delta x_{nt}^+ - H \Delta x_{nt} + A^T D_+ \vec{1} - A^T D \vec{1}) \\ &= \Delta x_{nt}^{+T} (-\nabla f_l(x^+) + \nabla f_l(x) + \nabla \phi(x^+) - \nabla \phi(x)) \\ &= 0\end{aligned}$$

Convergence of Newtons Method

$$\begin{aligned}DA\Delta x_{nt} &= DA(x^+ - x) \\ &= D(b - Ax - (b - Ax^+)) \\ &= D(D^{-1}\vec{1} - D_+^{-1}\vec{1}) \\ &= (I - D_+^{-1}D)\vec{1}\end{aligned}$$

$$a^T(a + b)$$

$$\begin{aligned}&= \Delta x_{nt}^{+T} A^T D_+ (D_+ A \Delta x_{nt}^+ + (I - D_+^{-1} D) DA \Delta x_{nt}) \\ &= \Delta x_{nt}^{+T} (A^T D_+^2 A \Delta x_{nt}^+ - A^T D^2 A \Delta x_{nt} + A^T D_+ DA \Delta x_{nt}) \\ &= \Delta x_{nt}^{+T} (H_+ \Delta x_{nt}^+ - H \Delta x_{nt} + A^T D_+ \vec{1} - A^T D \vec{1}) \\ &= \Delta x_{nt}^{+T} (-\nabla f_l(x^+) + \nabla f_l(x) + \nabla \phi(x^+) - \nabla \phi(x)) \\ &= 0\end{aligned}$$

Convergence of Newtons Method

$$\begin{aligned}DA\Delta x_{nt} &= DA(x^+ - x) \\ &= D(b - Ax - (b - Ax^+)) \\ &= D(D^{-1}\vec{1} - D_+^{-1}\vec{1}) \\ &= (I - D_+^{-1}D)\vec{1}\end{aligned}$$

$$a^T(a + b)$$

$$\begin{aligned}&= \Delta x_{nt}^{+T} A^T D_+ (D_+ A \Delta x_{nt}^+ + (I - D_+^{-1} D) DA \Delta x_{nt}) \\ &= \Delta x_{nt}^{+T} (A^T D_+^2 A \Delta x_{nt}^+ - A^T D^2 A \Delta x_{nt} + A^T D_+ DA \Delta x_{nt}) \\ &= \Delta x_{nt}^{+T} (H_+ \Delta x_{nt}^+ - H \Delta x_{nt} + A^T D_+ \vec{1} - A^T D \vec{1}) \\ &= \Delta x_{nt}^{+T} (-\nabla f_l(x^+) + \nabla f_l(x) + \nabla \phi(x^+) - \nabla \phi(x)) \\ &= 0\end{aligned}$$

Convergence of Newtons Method

$$\begin{aligned}DA\Delta x_{nt} &= DA(x^+ - x) \\ &= D(b - Ax - (b - Ax^+)) \\ &= D(D^{-1}\vec{1} - D_+^{-1}\vec{1}) \\ &= (I - D_+^{-1}D)\vec{1}\end{aligned}$$

$$a^T(a + b)$$

$$\begin{aligned}&= \Delta x_{nt}^{+T} A^T D_+ (D_+ A \Delta x_{nt}^+ + (I - D_+^{-1} D) DA \Delta x_{nt}) \\ &= \Delta x_{nt}^{+T} (A^T D_+^2 A \Delta x_{nt}^+ - A^T D^2 A \Delta x_{nt} + A^T D_+ DA \Delta x_{nt}) \\ &= \Delta x_{nt}^{+T} (H_+ \Delta x_{nt}^+ - H \Delta x_{nt} + A^T D_+ \vec{1} - A^T D \vec{1}) \\ &= \Delta x_{nt}^{+T} (-\nabla f_t(x^+) + \nabla f_t(x) + \nabla \phi(x^+) - \nabla \phi(x)) \\ &= 0\end{aligned}$$

Convergence of Newtons Method

$$\begin{aligned}DA\Delta x_{nt} &= DA(x^+ - x) \\ &= D(b - Ax - (b - Ax^+)) \\ &= D(D^{-1}\vec{1} - D_+^{-1}\vec{1}) \\ &= (I - D_+^{-1}D)\vec{1}\end{aligned}$$

$$\begin{aligned}a^T(a + b) &= \Delta x_{nt}^{+T} A^T D_+ \left(D_+ A \Delta x_{nt}^+ + (I - D_+^{-1}D) DA \Delta x_{nt} \right) \\ &= \Delta x_{nt}^{+T} \left(A^T D_+^2 A \Delta x_{nt}^+ - A^T D^2 A \Delta x_{nt} + A^T D_+ DA \Delta x_{nt} \right) \\ &= \Delta x_{nt}^{+T} \left(H_+ \Delta x_{nt}^+ - H \Delta x_{nt} + A^T D_+ \vec{1} - A^T D \vec{1} \right) \\ &= \Delta x_{nt}^{+T} \left(-\nabla f_t(x^+) + \nabla f_t(x) + \nabla \phi(x^+) - \nabla \phi(x) \right) \\ &= 0\end{aligned}$$

Convergence of Newtons Method

bound on $\lambda_t(\mathbf{x}^+)$:

we use $D := D_x = \text{diag}(d_x)$ and $D_+ := D_{x^+} = \text{diag}(d_{x^+})$

$$\begin{aligned}\lambda_t(\mathbf{x}^+)^2 &= \|D_+ A \Delta \mathbf{x}_{\text{nt}}^+\|^2 \\ &\leq \|D_+ A \Delta \mathbf{x}_{\text{nt}}^+\|^2 + \|D_+ A \Delta \mathbf{x}_{\text{nt}}^+ + (I - D_+^{-1} D) D A \Delta \mathbf{x}_{\text{nt}}\|^2 \\ &= \|(I - D_+^{-1} D) D A \Delta \mathbf{x}_{\text{nt}}\|^2 \\ &= \|(I - D_+^{-1} D)^2 \tilde{\mathbf{I}}\|^2 \\ &\leq \|(I - D_+^{-1} D) \tilde{\mathbf{I}}\|^4 \\ &= \|D A \Delta \mathbf{x}_{\text{nt}}\|^4 \\ &= \lambda_t(\mathbf{x})^4\end{aligned}$$

The second inequality follows from $\sum_i y_i^4 \leq (\sum_i y_i^2)^2$

Convergence of Newtons Method

bound on $\lambda_t(\mathbf{x}^+)$:

we use $D := D_x = \text{diag}(d_x)$ and $D_+ := D_{x^+} = \text{diag}(d_{x^+})$

$$\begin{aligned}\lambda_t(\mathbf{x}^+)^2 &= \|D_+ A \Delta \mathbf{x}_{\text{nt}}^+\|^2 \\ &\leq \|D_+ A \Delta \mathbf{x}_{\text{nt}}^+\|^2 + \|D_+ A \Delta \mathbf{x}_{\text{nt}}^+ + (I - D_+^{-1} D) D A \Delta \mathbf{x}_{\text{nt}}\|^2 \\ &= \|(I - D_+^{-1} D) D A \Delta \mathbf{x}_{\text{nt}}\|^2 \\ &= \|(I - D_+^{-1} D)^2 \vec{\Gamma}\|^2 \\ &\leq \|(I - D_+^{-1} D) \vec{\Gamma}\|^4 \\ &= \|D A \Delta \mathbf{x}_{\text{nt}}\|^4 \\ &= \lambda_t(\mathbf{x})^4\end{aligned}$$

The second inequality follows from $\sum_i y_i^4 \leq (\sum_i y_i^2)^2$

Convergence of Newtons Method

bound on $\lambda_t(\mathbf{x}^+)$:

we use $D := D_x = \text{diag}(d_x)$ and $D_+ := D_{x^+} = \text{diag}(d_{x^+})$

$$\begin{aligned}\lambda_t(\mathbf{x}^+)^2 &= \|D_+ A \Delta \mathbf{x}_{\text{nt}}^+\|^2 \\ &\leq \|D_+ A \Delta \mathbf{x}_{\text{nt}}^+\|^2 + \|D_+ A \Delta \mathbf{x}_{\text{nt}}^+ + (I - D_+^{-1} D) D A \Delta \mathbf{x}_{\text{nt}}\|^2 \\ &= \|(I - D_+^{-1} D) D A \Delta \mathbf{x}_{\text{nt}}\|^2 \\ &= \|(I - D_+^{-1} D)^2 \vec{\Gamma}\|^2 \\ &\leq \|(I - D_+^{-1} D) \vec{\Gamma}\|^4 \\ &= \|D A \Delta \mathbf{x}_{\text{nt}}\|^4 \\ &= \lambda_t(\mathbf{x})^4\end{aligned}$$

The second inequality follows from $\sum_i y_i^4 \leq (\sum_i y_i^2)^2$

Convergence of Newtons Method

bound on $\lambda_t(\mathbf{x}^+)$:

we use $D := D_x = \text{diag}(d_x)$ and $D_+ := D_{x^+} = \text{diag}(d_{x^+})$

$$\begin{aligned}\lambda_t(\mathbf{x}^+)^2 &= \|D_+ A \Delta \mathbf{x}_{\text{nt}}^+\|^2 \\ &\leq \|D_+ A \Delta \mathbf{x}_{\text{nt}}^+\|^2 + \|D_+ A \Delta \mathbf{x}_{\text{nt}}^+ + (I - D_+^{-1} D) D A \Delta \mathbf{x}_{\text{nt}}\|^2 \\ &= \|(I - D_+^{-1} D) D A \Delta \mathbf{x}_{\text{nt}}\|^2 \\ &= \|(I - D_+^{-1} D)^2 \vec{1}\|^2 \\ &\leq \|(I - D_+^{-1} D) \vec{1}\|^4 \\ &= \|D A \Delta \mathbf{x}_{\text{nt}}\|^4 \\ &= \lambda_t(\mathbf{x})^4\end{aligned}$$

The second inequality follows from $\sum_i y_i^4 \leq (\sum_i y_i^2)^2$

Convergence of Newtons Method

bound on $\lambda_t(\mathbf{x}^+)$:

we use $D := D_x = \text{diag}(d_x)$ and $D_+ := D_{x^+} = \text{diag}(d_{x^+})$

$$\begin{aligned}\lambda_t(\mathbf{x}^+)^2 &= \|D_+ A \Delta \mathbf{x}_{\text{nt}}^+\|^2 \\ &\leq \|D_+ A \Delta \mathbf{x}_{\text{nt}}^+\|^2 + \|D_+ A \Delta \mathbf{x}_{\text{nt}}^+ + (I - D_+^{-1} D) D A \Delta \mathbf{x}_{\text{nt}}\|^2 \\ &= \|(I - D_+^{-1} D) D A \Delta \mathbf{x}_{\text{nt}}\|^2 \\ &= \|(I - D_+^{-1} D)^2 \vec{\Gamma}\|^2 \\ &\leq \|(I - D_+^{-1} D) \vec{\Gamma}\|^4 \\ &= \|D A \Delta \mathbf{x}_{\text{nt}}\|^4 \\ &= \lambda_t(\mathbf{x})^4\end{aligned}$$

The second inequality follows from $\sum_i y_i^4 \leq (\sum_i y_i^2)^2$

If $\lambda_t(x)$ is large we do not have a guarantee.

Try to avoid this case!!!

Path-following Methods

Try to slowly travel along the central path.

Algorithm 1 PathFollowing

- 1: start at analytic center
- 2: **while** solution not good enough **do**
- 3: make step to improve objective function
- 4: recenter to return to central path

Short Step Barrier Method

simplifying assumptions:

- ▶ a first central point $x^*(t_0)$ is given
- ▶ $x^*(t)$ is computed exactly in each iteration

ϵ is approximation we are aiming for

start at $t = t_0$, repeat until $m/t \leq \epsilon$

- ▶ compute $x^*(\mu t)$ using Newton starting from $x^*(t)$
- ▶ $t := \mu t$

where $\mu = 1 + 1/(2\sqrt{m})$

Short Step Barrier Method

gradient of f_{t+} at $(x = x^*(t))$

$$\begin{aligned}\nabla f_{t+}(x) &= \nabla f_t(x) + (\mu - 1)tc \\ &= -(\mu - 1)A^T D_x \vec{1}\end{aligned}$$

This holds because $0 = \nabla f_t(x) = tc + A^T D_x \vec{1}$.

The Newton decrement is

$$\begin{aligned}\lambda_{t+}(x)^2 &= \nabla f_{t+}(x)^T H^{-1} \nabla f_{t+}(x) \\ &= (\mu - 1)^2 \vec{1}^T B (B^T B)^{-1} B^T \vec{1} \quad B = D_x^T A \\ &\leq (\mu - 1)^2 m \\ &= 1/4\end{aligned}$$

This means we are in the range of quadratic convergence!!!

Short Step Barrier Method

gradient of f_{t+} at $(x = x^*(t))$

$$\begin{aligned}\nabla f_{t+}(x) &= \nabla f_t(x) + (\mu - 1)tc \\ &= -(\mu - 1)A^T D_x \vec{1}\end{aligned}$$

This holds because $0 = \nabla f_t(x) = tc + A^T D_x \vec{1}$.

The Newton decrement is

$$\begin{aligned}\lambda_{t+}(x)^2 &= \nabla f_{t+}(x)^T H^{-1} \nabla f_{t+}(x) \\ &= (\mu - 1)^2 \vec{1}^T B (B^T B)^{-1} B^T \vec{1} \quad B = D_x^T A \\ &\leq (\mu - 1)^2 m \\ &= 1/4\end{aligned}$$

This means we are in the range of quadratic convergence!!!

Short Step Barrier Method

gradient of f_{t+} at $(x = x^*(t))$

$$\begin{aligned}\nabla f_{t+}(x) &= \nabla f_t(x) + (\mu - 1)tc \\ &= -(\mu - 1)A^T D_x \vec{1}\end{aligned}$$

This holds because $0 = \nabla f_t(x) = tc + A^T D_x \vec{1}$.

The Newton decrement is

$$\begin{aligned}\lambda_{t+}(x)^2 &= \nabla f_{t+}(x)^T H^{-1} \nabla f_{t+}(x) \\ &= (\mu - 1)^2 \vec{1}^T B (B^T B)^{-1} B^T \vec{1} \quad B = D_x^T A \\ &\leq (\mu - 1)^2 m \\ &= 1/4\end{aligned}$$

This means we are in the range of quadratic convergence!!!

Short Step Barrier Method

gradient of f_{t+} at $(x = x^*(t))$

$$\begin{aligned}\nabla f_{t+}(x) &= \nabla f_t(x) + (\mu - 1)tc \\ &= -(\mu - 1)A^T D_x \vec{1}\end{aligned}$$

This holds because $0 = \nabla f_t(x) = tc + A^T D_x \vec{1}$.

The Newton decrement is

$$\begin{aligned}\lambda_{t+}(x)^2 &= \nabla f_{t+}(x)^T H^{-1} \nabla f_{t+}(x) \\ &= (\mu - 1)^2 \vec{1}^T B (B^T B)^{-1} B^T \vec{1} \quad B = D_x^T A \\ &\leq (\mu - 1)^2 m \\ &= 1/4\end{aligned}$$

This means we are in the range of quadratic convergence!!!

Short Step Barrier Method

gradient of f_{t+} at $(x = x^*(t))$

$$\begin{aligned}\nabla f_{t+}(x) &= \nabla f_t(x) + (\mu - 1)tc \\ &= -(\mu - 1)A^T D_x \vec{1}\end{aligned}$$

This holds because $0 = \nabla f_t(x) = tc + A^T D_x \vec{1}$.

The Newton decrement is

$$\begin{aligned}\lambda_{t+}(x)^2 &= \nabla f_{t+}(x)^T H^{-1} \nabla f_{t+}(x) \\ &= (\mu - 1)^2 \vec{1}^T B (B^T B)^{-1} B^T \vec{1} \quad B = D_x^T A \\ &\leq (\mu - 1)^2 m \\ &= 1/4\end{aligned}$$

This means we are in the range of quadratic convergence!!!

Number of Iterations

the number of Newton iterations per outer iteration is very small; in practise only 1 or 2

Number of outer iterations:

We need $t_k = \mu^k t_0 \geq m/\epsilon$. This holds when

$$k \geq \frac{\log(m/(\epsilon t_0))}{\log(\mu)}$$

We get a bound of

$$\mathcal{O}\left(\sqrt{m} \log \frac{m}{\epsilon t_0}\right)$$

We show how to get a starting point with $t_0 = 1/2^L$. Together with $\epsilon \approx 2^{-L}$ we get $\mathcal{O}(L\sqrt{m})$ iterations.

Explanation for previous slide

$P = B(B^T B)^{-1} B^T$ is a symmetric real-valued matrix; it has n linearly independent Eigenvectors. Since it is a **projection matrix** ($P^2 = P$) it can only have Eigenvalues 0 and 1 (because the Eigenvalues of P^2 are λ_i^2 , where λ_i is Eigenvalue of P).

The expression

$$\max_v \frac{v^T P v}{v^T v}$$

gives the largest Eigenvalue for P . Hence, $\vec{1}^T P \vec{1} \leq \vec{1}^T \vec{1} = m$

Damped Newton Method

We assume that the polytope (not just the LP) is bounded. Then $Av \leq 0$ is not possible.

For $x \in P^\circ$ and direction $v \neq 0$ define

$$\sigma_x(v) := \max_i \frac{a_i^T v}{s_i(x)}$$

$a_i^T v$ is the change on the left hand side of the i -th constraint when moving in direction of v .

If $\sigma_x(v) > 1$ then for one coordinate this change is larger than the slack in the constraint at position x .

By downscaling v we can ensure to stay in the polytope.

Observation:

$$x + \alpha v \in P \quad \text{for } \alpha \in \{0, 1/\sigma_x(v)\}$$

Damped Newton Method

Suppose that we move from x to $x + \alpha v$. The linear estimate says that $f_t(x)$ should change by $\nabla f_t(x)^T \alpha v$.

The following argument shows that f_t is well behaved. For small α the reduction of $f_t(x)$ is close to linear estimate.

$$f_t(x + \alpha v) - f_t(x) = t c^T \alpha v + \phi(x + \alpha v) - \phi(x)$$

$$\phi(x + \alpha v) - \phi(x)$$

$$s_i(x + \alpha v) = b_i - a_i^T x - a_i^T \alpha v = s_i(x) - a_i^T \alpha v$$

Damped Newton Method

Suppose that we move from x to $x + \alpha v$. The linear estimate says that $f_t(x)$ should change by $\nabla f_t(x)^T \alpha v$.

The following argument shows that f_t is well behaved. For small α the reduction of $f_t(x)$ is close to linear estimate.

$$f_t(x + \alpha v) - f_t(x) = tc^T \alpha v + \phi(x + \alpha v) - \phi(x)$$

$$\phi(x + \alpha v) - \phi(x)$$

$$s_i(x + \alpha v) = b_i - a_i^T x - a_i^T \alpha v = s_i(x) - a_i^T \alpha v$$

Damped Newton Method

Suppose that we move from x to $x + \alpha v$. The linear estimate says that $f_t(x)$ should change by $\nabla f_t(x)^T \alpha v$.

The following argument shows that f_t is well behaved. For small α the reduction of $f_t(x)$ is close to linear estimate.

$$f_t(x + \alpha v) - f_t(x) = t c^T \alpha v + \phi(x + \alpha v) - \phi(x)$$

$$\begin{aligned} \phi(x + \alpha v) - \phi(x) &= - \sum_i \log(s_i(x + \alpha v)) + \sum_i \log(s_i(x)) \\ &= - \sum_i \log(s_i(x + \alpha v) / s_i(x)) \\ &= - \sum_i \log(1 - a_i^T \alpha v / s_i(x)) \end{aligned}$$

$$s_i(x + \alpha v) = b_i - a_i^T x - a_i^T \alpha v = s_i(x) - a_i^T \alpha v$$

Damped Newton Method

Suppose that we move from x to $x + \alpha v$. The linear estimate says that $f_t(x)$ should change by $\nabla f_t(x)^T \alpha v$.

The following argument shows that f_t is well behaved. For small α the reduction of $f_t(x)$ is close to linear estimate.

$$f_t(x + \alpha v) - f_t(x) = t c^T \alpha v + \phi(x + \alpha v) - \phi(x)$$

$$\begin{aligned} \phi(x + \alpha v) - \phi(x) &= - \sum_i \log(s_i(x + \alpha v)) + \sum_i \log(s_i(x)) \\ &= - \sum_i \log(s_i(x + \alpha v)/s_i(x)) \\ &= - \sum_i \log(1 - a_i^T \alpha v / s_i(x)) \end{aligned}$$

$$s_i(x + \alpha v) = b_i - a_i^T x - a_i^T \alpha v = s_i(x) - a_i^T \alpha v$$

Damped Newton Method

Suppose that we move from x to $x + \alpha v$. The linear estimate says that $f_t(x)$ should change by $\nabla f_t(x)^T \alpha v$.

The following argument shows that f_t is well behaved. For small α the reduction of $f_t(x)$ is close to linear estimate.

$$f_t(x + \alpha v) - f_t(x) = t c^T \alpha v + \phi(x + \alpha v) - \phi(x)$$

$$\begin{aligned}\phi(x + \alpha v) - \phi(x) &= -\sum_i \log(s_i(x + \alpha v)) + \sum_i \log(s_i(x)) \\ &= -\sum_i \log(s_i(x + \alpha v)/s_i(x)) \\ &= -\sum_i \log(1 - a_i^T \alpha v / s_i(x))\end{aligned}$$

$$s_i(x + \alpha v) = b_i - a_i^T x - a_i^T \alpha v = s_i(x) - a_i^T \alpha v$$

Damped Newton Method

$$\begin{aligned}\nabla f_t(x)^T \alpha v &= (tc^T + \sum_i a_i^T / s_i(x)) \alpha v \\ &= tc^T \alpha v + \sum_i \alpha w_i\end{aligned}$$

Note that $\|w\| = \|v\|_{H_x}$.

Define $w_i = a_i^T v / s_i(x)$ and $\sigma = \max_i w_i$. Then

$$f_t(x + \alpha v) - f_t(x) - \nabla f_t(x)^T \alpha v$$

For $|x| < 1$, $x \leq 0$:

$$x + \log(1 - x) = -\frac{x^2}{2} - \frac{x^3}{3} - \frac{x^4}{4} - \dots \geq -\frac{x^2}{2} = -\frac{y^2}{2} \frac{x^2}{y^2}$$

For $|x| < 1$, $0 < x \leq y$:

$$\begin{aligned}x + \log(1 - x) &= -\frac{x^2}{2} - \frac{x^3}{3} - \frac{x^4}{4} - \dots = \frac{x^2}{y^2} \left(-\frac{y^2}{2} - \frac{y^2 x}{3} - \frac{y^2 x^2}{4} - \dots \right) \\ &\geq \frac{x^2}{y^2} \left(-\frac{y^2}{2} - \frac{y^3}{3} - \frac{y^4}{4} - \dots \right) = \frac{x^2}{y^2} (y + \log(1 - y))\end{aligned}$$

Damped Newton Method

$$\begin{aligned}\nabla f_t(x)^T \alpha v &= (tc^T + \sum_i a_i^T / s_i(x)) \alpha v \\ &= tc^T \alpha v + \sum_i \alpha w_i\end{aligned}$$

Define $w_i = a_i^T v / s_i(x)$ and $\sigma = \max_i w_i$. Then

Note that $\|w\| = \|v\|_{H_x}$.

$$\begin{aligned}f_t(x + \alpha v) - f_t(x) - \nabla f_t(x)^T \alpha v &= - \sum_i (\alpha w_i + \log(1 - \alpha w_i)) \\ &\leq - \sum_{w_i > 0} (\alpha w_i + \log(1 - \alpha w_i)) + \sum_{w_i \leq 0} \frac{\alpha^2 w_i^2}{2} \\ &\leq - \sum_{w_i > 0} \frac{w_i^2}{\sigma^2} (\alpha \sigma + \log(1 - \alpha \sigma)) + \frac{(\alpha \sigma)^2}{2} \sum_{w_i \leq 0} \frac{w_i^2}{\sigma^2}\end{aligned}$$

For $|x| < 1, x \leq 0$:

$$x + \log(1 - x) = -\frac{x^2}{2} - \frac{x^3}{3} - \frac{x^4}{4} - \dots \geq -\frac{x^2}{2} = -\frac{y^2}{2} \frac{x^2}{y^2}$$

For $|x| < 1, 0 < x \leq y$:

$$\begin{aligned}x + \log(1 - x) &= -\frac{x^2}{2} - \frac{x^3}{3} - \frac{x^4}{4} - \dots = \frac{x^2}{y^2} \left(-\frac{y^2}{2} - \frac{y^2 x}{3} - \frac{y^2 x^2}{4} - \dots \right) \\ &\geq \frac{x^2}{y^2} \left(-\frac{y^2}{2} - \frac{y^3}{3} - \frac{y^4}{4} - \dots \right) = \frac{x^2}{y^2} (y + \log(1 - y))\end{aligned}$$

Damped Newton Method

$$\begin{aligned}\nabla f_t(x)^T \alpha v &= (t c^T + \sum_i a_i^T / s_i(x)) \alpha v \\ &= t c^T \alpha v + \sum_i \alpha w_i\end{aligned}$$

Define $w_i = a_i^T v / s_i(x)$ and $\sigma = \max_i w_i$. Then

Note that $\|w\| = \|v\|_{H_x}$.

$$\begin{aligned}f_t(x + \alpha v) - f_t(x) - \nabla f_t(x)^T \alpha v &= - \sum_i (\alpha w_i + \log(1 - \alpha w_i)) \\ &\leq - \sum_{w_i > 0} (\alpha w_i + \log(1 - \alpha w_i)) + \sum_{w_i \leq 0} \frac{\alpha^2 w_i^2}{2} \\ &\leq - \sum_{w_i > 0} \frac{w_i^2}{\sigma^2} (\alpha \sigma + \log(1 - \alpha \sigma)) + \frac{(\alpha \sigma)^2}{2} \sum_{w_i \leq 0} \frac{w_i^2}{\sigma^2}\end{aligned}$$

For $|x| < 1$, $x \leq 0$:

$$x + \log(1 - x) = -\frac{x^2}{2} - \frac{x^3}{3} - \frac{x^4}{4} - \dots \geq -\frac{x^2}{2} = -\frac{y^2}{2} \frac{x^2}{y^2}$$

For $|x| < 1$, $0 < x \leq y$:

$$\begin{aligned}x + \log(1 - x) &= -\frac{x^2}{2} - \frac{x^3}{3} - \frac{x^4}{4} - \dots = \frac{x^2}{y^2} \left(-\frac{y^2}{2} - \frac{y^2 x}{3} - \frac{y^2 x^2}{4} - \dots \right) \\ &\geq \frac{x^2}{y^2} \left(-\frac{y^2}{2} - \frac{y^3}{3} - \frac{y^4}{4} - \dots \right) = \frac{x^2}{y^2} (y + \log(1 - y))\end{aligned}$$

Damped Newton Method

$$\begin{aligned}\nabla f_t(x)^T \alpha v &= (tc^T + \sum_i a_i^T / s_i(x)) \alpha v \\ &= tc^T \alpha v + \sum_i \alpha w_i\end{aligned}$$

Define $w_i = a_i^T v / s_i(x)$ and $\sigma = \max_i w_i$. Then

Note that $\|w\| = \|v\|_{H_x}$.

$$\begin{aligned}f_t(x + \alpha v) - f_t(x) - \nabla f_t(x)^T \alpha v &= - \sum_i (\alpha w_i + \log(1 - \alpha w_i)) \\ &\leq - \sum_{w_i > 0} (\alpha w_i + \log(1 - \alpha w_i)) + \sum_{w_i \leq 0} \frac{\alpha^2 w_i^2}{2} \\ &\leq - \sum_{w_i > 0} \frac{w_i^2}{\sigma^2} (\alpha \sigma + \log(1 - \alpha \sigma)) + \frac{(\alpha \sigma)^2}{2} \sum_{w_i \leq 0} \frac{w_i^2}{\sigma^2}\end{aligned}$$

For $|x| < 1, x \leq 0$:

$$x + \log(1 - x) = -\frac{x^2}{2} - \frac{x^3}{3} - \frac{x^4}{4} - \dots \geq -\frac{x^2}{2} = -\frac{y^2}{2} \frac{x^2}{y^2}$$

For $|x| < 1, 0 < x \leq y$:

$$\begin{aligned}x + \log(1 - x) &= -\frac{x^2}{2} - \frac{x^3}{3} - \frac{x^4}{4} - \dots = \frac{x^2}{y^2} \left(-\frac{y^2}{2} - \frac{y^2 x}{3} - \frac{y^2 x^2}{4} - \dots \right) \\ &\geq \frac{x^2}{y^2} \left(-\frac{y^2}{2} - \frac{y^3}{3} - \frac{y^4}{4} - \dots \right) = \frac{x^2}{y^2} (y + \log(1 - y))\end{aligned}$$

Damped Newton Method

$$\text{For } x \geq 0 \quad \frac{x^2}{2} \leq \frac{x^2}{2} + \frac{x^3}{3} + \frac{x^4}{4} + \dots = -(x + \log(1 - x))$$

$$\begin{aligned} &\leq -\sum_i \frac{w_i^2}{\sigma^2} (\alpha\sigma + \log(1 - \alpha\sigma)) \\ &= -\frac{1}{\sigma^2} \|v\|_{H_x}^2 (\alpha\sigma + \log(1 - \alpha\sigma)) \end{aligned}$$

Damped Newton Iteration:

In a damped Newton step we choose

$$x_+ = x + \frac{1}{1 + \sigma_x(\Delta x_{nt})} \Delta x_{nt}$$

This means that in the above expressions we choose $\alpha = \frac{1}{1 + \sigma}$ and $v = \Delta x_{nt}$. Note that it wouldn't make sense to choose α larger than 1 as this would mean that our real target ($x + \Delta x_{nt}$) is inside the polytope but we overshoot and go further than this target.

Damped Newton Method

$$\text{For } x \geq 0 \quad \frac{x^2}{2} \leq \frac{x^2}{2} + \frac{x^3}{3} + \frac{x^4}{4} + \dots = -(x + \log(1-x))$$

$$\begin{aligned} &\leq -\sum_i \frac{w_i^2}{\sigma^2} (\alpha\sigma + \log(1-\alpha\sigma)) \\ &= -\frac{1}{\sigma^2} \|v\|_{H_x}^2 (\alpha\sigma + \log(1-\alpha\sigma)) \end{aligned}$$

Damped Newton Iteration:

In a damped Newton step we choose

$$x_+ = x + \frac{1}{1 + \sigma_x(\Delta x_{nt})} \Delta x_{nt}$$

This means that in the above expressions we choose $\alpha = \frac{1}{1+\sigma}$ and $v = \Delta x_{nt}$. Note that it wouldn't make sense to choose α larger than 1 as this would mean that our real target ($x + \Delta x_{nt}$) is inside the polytope but we overshoot and go further than this target.

Damped Newton Method

$$\text{For } x \geq 0 \quad \frac{x^2}{2} \leq \frac{x^2}{2} + \frac{x^3}{3} + \frac{x^4}{4} + \dots = -(x + \log(1-x))$$

$$\begin{aligned} &\leq -\sum_i \frac{w_i^2}{\sigma^2} (\alpha\sigma + \log(1-\alpha\sigma)) \\ &= -\frac{1}{\sigma^2} \|v\|_{H_x}^2 (\alpha\sigma + \log(1-\alpha\sigma)) \end{aligned}$$

Damped Newton Iteration:

In a damped Newton step we choose

$$x_+ = x + \frac{1}{1 + \sigma_x(\Delta x_{nt})} \Delta x_{nt}$$

This means that in the above expressions we choose $\alpha = \frac{1}{1+\sigma}$ and $v = \Delta x_{nt}$. Note that it wouldn't make sense to choose α larger than 1 as this would mean that our real target $(x + \Delta x_{nt})$ is inside the polytope but we overshoot and go further than this target.

Damped Newton Method

Theorem:

In a damped Newton step the cost decreases by at least

$$\lambda_t(x) - \log(1 + \lambda_t(x))$$

Proof: The decrease in cost is

$$-\alpha \nabla f_t(x)^T v + \frac{1}{\sigma^2} \|v\|_{H_x}^2 (\alpha\sigma + \log(1 - \alpha\sigma))$$

Choosing $\alpha = \frac{1}{1+\sigma}$ and $v = \Delta x_{nt}$ gives

With $v = \Delta x_{nt}$ we have $\|w\|_2 = \|v\|_{H_x} = \lambda_t(x)$; further recall that $\sigma = \|w\|_\infty$; hence $\sigma \leq \lambda_t(x)$.

Damped Newton Method

Theorem:

In a damped Newton step the cost decreases by at least

$$\lambda_t(x) - \log(1 + \lambda_t(x))$$

Proof: The decrease in cost is

$$-\alpha \nabla f_t(x)^T v + \frac{1}{\sigma^2} \|v\|_{H_x}^2 (\alpha\sigma + \log(1 - \alpha\sigma))$$

Choosing $\alpha = \frac{1}{1+\sigma}$ and $v = \Delta x_{nt}$ gives

With $v = \Delta x_{nt}$ we have $\|w\|_2 = \|v\|_{H_x} = \lambda_t(x)$; further recall that $\sigma = \|w\|_\infty$; hence $\sigma \leq \lambda_t(x)$.

Damped Newton Method

Theorem:

In a damped Newton step the cost decreases by at least

$$\lambda_t(x) - \log(1 + \lambda_t(x))$$

Proof: The decrease in cost is

$$-\alpha \nabla f_t(x)^T v + \frac{1}{\sigma^2} \|v\|_{H_x}^2 (\alpha \sigma + \log(1 - \alpha \sigma))$$

Choosing $\alpha = \frac{1}{1+\sigma}$ and $v = \Delta x_{nt}$ gives

$$\begin{aligned} \frac{1}{1+\sigma} \lambda_t(x)^2 + \frac{\lambda_t(x)^2}{\sigma^2} \left(\frac{\sigma}{1+\sigma} + \log \left(1 - \frac{\sigma}{1+\sigma} \right) \right) \\ = \frac{\lambda_t(x)^2}{\sigma^2} (\sigma - \log(1 + \sigma)) \end{aligned}$$

With $v = \Delta x_{nt}$ we have $\|w\|_2 = \|v\|_{H_x} = \lambda_t(x)$; further recall that $\sigma = \|w\|_\infty$; hence $\sigma \leq \lambda_t(x)$.

Damped Newton Method

Theorem:

In a damped Newton step the cost decreases by at least

$$\lambda_t(x) - \log(1 + \lambda_t(x))$$

Proof: The decrease in cost is

$$-\alpha \nabla f_t(x)^T v + \frac{1}{\sigma^2} \|v\|_{H_x}^2 (\alpha \sigma + \log(1 - \alpha \sigma))$$

Choosing $\alpha = \frac{1}{1+\sigma}$ and $v = \Delta x_{nt}$ gives

$$\begin{aligned} \frac{1}{1+\sigma} \lambda_t(x)^2 + \frac{\lambda_t(x)^2}{\sigma^2} \left(\frac{\sigma}{1+\sigma} + \log \left(1 - \frac{\sigma}{1+\sigma} \right) \right) \\ = \frac{\lambda_t(x)^2}{\sigma^2} (\sigma - \log(1 + \sigma)) \end{aligned}$$

With $v = \Delta x_{nt}$ we have $\|w\|_2 = \|v\|_{H_x} = \lambda_t(x)$; further recall that $\sigma = \|w\|_\infty$; hence $\sigma \leq \lambda_t(x)$.

Damped Newton Method

The first inequality follows since the function $\frac{1}{x^2}(x - \log(1+x))$ is monotonically decreasing.

$$\begin{aligned} &\geq \lambda_t(\mathbf{x}) - \log(1 + \lambda_t(\mathbf{x})) \\ &\geq 0.09 \end{aligned}$$

for $\lambda_t(\mathbf{x}) \geq 0.5$

Centering Algorithm:

Input: precision δ ; starting point x

1. compute Δx_{nt} and $\lambda_t(x)$
2. if $\lambda_t(x) \leq \delta$ return x
3. set $x := x + \alpha \Delta x_{nt}$ with

$$\alpha = \begin{cases} \frac{1}{1 + \sigma_x(\Delta x_{nt})} & \lambda_t \geq 1/2 \\ 1 & \text{otw.} \end{cases}$$

Damped Newton Method

The first inequality follows since the function $\frac{1}{x^2}(x - \log(1+x))$ is monotonically decreasing.

$$\begin{aligned} &\geq \lambda_t(x) - \log(1 + \lambda_t(x)) \\ &\geq 0.09 \end{aligned}$$

for $\lambda_t(x) \geq 0.5$

Centering Algorithm:

Input: precision δ ; starting point x

1. compute Δx_{nt} and $\lambda_t(x)$
2. if $\lambda_t(x) \leq \delta$ return x
3. set $x := x + \alpha \Delta x_{nt}$ with

$$\alpha = \begin{cases} \frac{1}{1 + \sigma_x(\Delta x_{nt})} & \lambda_t \geq 1/2 \\ 1 & \text{otw.} \end{cases}$$

Centering

Lemma 3

The centering algorithm starting at x_0 reaches a point with $\lambda_t(x) \leq \delta$ after

$$\frac{f_t(x_0) - \min_y f_t(y)}{0.09} + \mathcal{O}(\log \log(1/\delta))$$

iterations.

This can be very, very slow...

How to get close to analytic center?

Let $P = \{Ax \leq b\}$ be our (**feasible**) polyhedron, and x_0 a feasible point.

We change $b \rightarrow b + \frac{1}{\lambda} \cdot \vec{1}$, where $L = \langle A \rangle + \langle b \rangle + \langle c \rangle$ (**encoding length**) and $\lambda = 2^{2L}$. Recall that a basis is feasible in the old LP iff it is feasible in the new LP.

How to get close to analytic center?

Let $P = \{Ax \leq b\}$ be our (**feasible**) polyhedron, and x_0 a feasible point.

We change $b \rightarrow b + \frac{1}{\lambda} \cdot \vec{1}$, where $L = \langle A \rangle + \langle b \rangle + \langle c \rangle$ (**encoding length**) and $\lambda = 2^{2L}$. Recall that a basis is feasible in the old LP iff it is feasible in the new LP.

Lemma [without proof]

The inverse of a matrix M can be represented with rational numbers that have denominators $z_{ij} = \det(M)$.

For two basis solutions $x_B, x_{\bar{B}}$, the cost-difference $c^T x_B - c^T x_{\bar{B}}$ can be represented by a rational number that has denominator $z = \det(A_B) \cdot \det(A_{\bar{B}})$.

This means that in the perturbed LP it is sufficient to decrease the duality gap to $1/2^{4L}$ (i.e., $t \approx 2^{4L}$). This means the previous analysis essentially also works for the perturbed LP.

For a point x from the polytope (not necessarily BFS) the objective value $\bar{c}^T x$ is at most $n2^M 2^L$, where $M \leq L$ is the encoding length of the largest entry in \bar{c} .

Lemma [without proof]

The inverse of a matrix M can be represented with rational numbers that have denominators $z_{ij} = \det(M)$.

For two basis solutions $x_B, x_{\bar{B}}$, the cost-difference $c^T x_B - c^T x_{\bar{B}}$ can be represented by a rational number that has denominator $z = \det(A_B) \cdot \det(A_{\bar{B}})$.

This means that in the perturbed LP it is sufficient to decrease the duality gap to $1/2^{4L}$ (i.e., $t \approx 2^{4L}$). This means the previous analysis essentially also works for the perturbed LP.

For a point x from the polytope (not necessarily BFS) the objective value $\bar{c}^T x$ is at most $n2^M 2^L$, where $M \leq L$ is the encoding length of the largest entry in \bar{c} .

Lemma [without proof]

The inverse of a matrix M can be represented with rational numbers that have denominators $z_{ij} = \det(M)$.

For two basis solutions $x_B, x_{\bar{B}}$, the cost-difference $c^T x_B - c^T x_{\bar{B}}$ can be represented by a rational number that has denominator $z = \det(A_B) \cdot \det(A_{\bar{B}})$.

This means that in the perturbed LP it is sufficient to decrease the duality gap to $1/2^{4L}$ (i.e., $t \approx 2^{4L}$). This means the previous analysis essentially also works for the perturbed LP.

For a point x from the polytope (not necessarily BFS) the objective value $\bar{c}^T x$ is at most $n2^M 2^L$, where $M \leq L$ is the encoding length of the largest entry in \bar{c} .

Lemma [without proof]

The inverse of a matrix M can be represented with rational numbers that have denominators $z_{ij} = \det(M)$.

For two basis solutions $x_B, x_{\bar{B}}$, the cost-difference $c^T x_B - c^T x_{\bar{B}}$ can be represented by a rational number that has denominator $z = \det(A_B) \cdot \det(A_{\bar{B}})$.

This means that in the perturbed LP it is sufficient to decrease the duality gap to $1/2^{4L}$ (i.e., $t \approx 2^{4L}$). This means the previous analysis essentially also works for the perturbed LP.

For a point x from the polytope (not necessarily BFS) the objective value $\bar{c}^T x$ is at most $n2^M 2^L$, where $M \leq L$ is the encoding length of the largest entry in \bar{c} .

How to get close to analytic center?

Start at x_0 .

Choose $\hat{c} := -\nabla \phi(x)$.

$x_0 = x^*(1)$ is point on central path for \hat{c} and $t = 1$.

You can travel the central path in both directions. Go towards 0 until $t \approx 1/2^{\Omega(L)}$. This requires $O(\sqrt{m}L)$ outer iterations.

Let $x_{\hat{c}}$ denote this point.

Let x_c denote the point that minimizes

$$t \cdot c^T x + \phi(x)$$

(i.e., same value for t but different c , hence, different central path).

Note that an entry in \hat{c} fulfills $|\hat{c}_i| \leq 2^{2L}$. This holds since the slack in every constraint at x_0 is at least $\lambda = 1/2^{2L}$, and the gradient is the vector of inverse slacks.

How to get close to analytic center?

Start at x_0 .

Choose $\hat{c} := -\nabla\phi(x)$.

Note that an entry in \hat{c} fulfills $|\hat{c}_i| \leq 2^{2L}$. This holds since the slack in every constraint at x_0 is at least $\lambda = 1/2^{2L}$, and the gradient is the vector of inverse slacks.

$x_0 = x^*(1)$ is point on central path for \hat{c} and $t = 1$.

You can travel the central path in both directions. Go towards 0 until $t \approx 1/2^{\Omega(L)}$. This requires $O(\sqrt{m}L)$ outer iterations.

Let $x_{\hat{c}}$ denote this point.

Let x_c denote the point that minimizes

$$t \cdot c^T x + \phi(x)$$

(i.e., same value for t but different c , hence, different central path).

How to get close to analytic center?

Start at x_0 .

Choose $\hat{c} := -\nabla\phi(x)$.

Note that an entry in \hat{c} fulfills $|\hat{c}_i| \leq 2^{2L}$.
This holds since the slack in every constraint at x_0 is at least $\lambda = 1/2^{2L}$, and the gradient is the vector of inverse slacks.

$x_0 = x^*(1)$ is point on central path for \hat{c} and $t = 1$.

You can travel the central path in both directions. Go towards 0 until $t \approx 1/2^{\Omega(L)}$. This requires $O(\sqrt{m}L)$ outer iterations.

Let $x_{\hat{c}}$ denote this point.

Let x_c denote the point that minimizes

$$t \cdot c^T x + \phi(x)$$

(i.e., same value for t but different c , hence, different central path).

How to get close to analytic center?

Start at x_0 .

Choose $\hat{c} := -\nabla\phi(x)$.

Note that an entry in \hat{c} fulfills $|\hat{c}_i| \leq 2^{2L}$. This holds since the slack in every constraint at x_0 is at least $\lambda = 1/2^{2L}$, and the gradient is the vector of inverse slacks.

$x_0 = x^*(1)$ is point on central path for \hat{c} and $t = 1$.

You can travel the central path in both directions. Go towards 0 until $t \approx 1/2^{\Omega(L)}$. This requires $O(\sqrt{m}L)$ outer iterations.

Let $x_{\hat{c}}$ denote this point.

Let x_c denote the point that minimizes

$$t \cdot c^T x + \phi(x)$$

(i.e., same value for t but different c , hence, different central path).

How to get close to analytic center?

Start at x_0 .

Choose $\hat{c} := -\nabla\phi(x)$.

Note that an entry in \hat{c} fulfills $|\hat{c}_i| \leq 2^{2L}$. This holds since the slack in every constraint at x_0 is at least $\lambda = 1/2^{2L}$, and the gradient is the vector of inverse slacks.

$x_0 = x^*(1)$ is point on central path for \hat{c} and $t = 1$.

You can travel the central path in both directions. Go towards 0 until $t \approx 1/2^{\Omega(L)}$. This requires $O(\sqrt{m}L)$ outer iterations.

Let $x_{\hat{c}}$ denote this point.

Let x_c denote the point that minimizes

$$t \cdot c^T x + \phi(x)$$

(i.e., same value for t but different c , hence, different central path).

How to get close to analytic center?

Start at x_0 .

Choose $\hat{c} := -\nabla \phi(x)$.

Note that an entry in \hat{c} fulfills $|\hat{c}_i| \leq 2^{2L}$. This holds since the slack in every constraint at x_0 is at least $\lambda = 1/2^{2L}$, and the gradient is the vector of inverse slacks.

$x_0 = x^*(1)$ is point on central path for \hat{c} and $t = 1$.

You can travel the central path in both directions. Go towards 0 until $t \approx 1/2^{\Omega(L)}$. This requires $O(\sqrt{m}L)$ outer iterations.

Let $x_{\hat{c}}$ denote this point.

Let x_c denote the point that minimizes

$$t \cdot c^T x + \phi(x)$$

(i.e., same value for t but different c , hence, different central path).

How to get close to analytic center?

Clearly,

$$t \cdot \hat{c}^T x_{\hat{c}} + \phi(x_{\hat{c}}) \leq t \cdot \hat{c}^T x_c + \phi(x_c)$$

The difference between $f_t(x_{\hat{c}})$ and $f_t(x_c)$ is

$$\begin{aligned} t c^T x_{\hat{c}} + \phi(x_{\hat{c}}) - t c^T x_c - \phi(x_c) \\ \leq t(c^T x_{\hat{c}} + \hat{c}^T x_c - \hat{c}^T x_{\hat{c}} - c^T x_c) \\ \leq 4tn2^{3L} \end{aligned}$$

For $t = 1/2^{\Omega(L)}$ the last term becomes constant. Hence, using damped Newton we can move from $x_{\hat{c}}$ to x_c quickly.

In total for this analysis we require $\mathcal{O}(\sqrt{m}L)$ outer iterations for the whole algorithm.

One iteration can be implemented in $\tilde{O}(m^3)$ time.

How to get close to analytic center?

Clearly,

$$t \cdot \hat{c}^T x_{\hat{c}} + \phi(x_{\hat{c}}) \leq t \cdot \hat{c}^T x_c + \phi(x_c)$$

The difference between $f_t(x_{\hat{c}})$ and $f_t(x_c)$ is

$$\begin{aligned} t c^T x_{\hat{c}} + \phi(x_{\hat{c}}) - t c^T x_c - \phi(x_c) \\ \leq t(c^T x_{\hat{c}} + \hat{c}^T x_c - \hat{c}^T x_{\hat{c}} - c^T x_c) \\ \leq 4tn2^{3L} \end{aligned}$$

For $t = 1/2^{\Omega(L)}$ the last term becomes constant. Hence, using damped Newton we can move from $x_{\hat{c}}$ to x_c quickly.

In total for this analysis we require $\mathcal{O}(\sqrt{m}L)$ outer iterations for the whole algorithm.

One iteration can be implemented in $\tilde{\mathcal{O}}(m^3)$ time.

How to get close to analytic center?

Clearly,

$$t \cdot \hat{c}^T x_{\hat{c}} + \phi(x_{\hat{c}}) \leq t \cdot \hat{c}^T x_c + \phi(x_c)$$

The difference between $f_t(x_{\hat{c}})$ and $f_t(x_c)$ is

$$\begin{aligned} t c^T x_{\hat{c}} + \phi(x_{\hat{c}}) - t c^T x_c - \phi(x_c) \\ \leq t(c^T x_{\hat{c}} + \hat{c}^T x_c - \hat{c}^T x_{\hat{c}} - c^T x_c) \\ \leq 4tn2^{3L} \end{aligned}$$

For $t = 1/2^{\Omega(L)}$ the last term becomes constant. Hence, using damped Newton we can move from $x_{\hat{c}}$ to x_c quickly.

In total for this analysis we require $\mathcal{O}(\sqrt{m}L)$ outer iterations for the whole algorithm.

One iteration can be implemented in $\tilde{\mathcal{O}}(m^3)$ time.

How to get close to analytic center?

Clearly,

$$t \cdot \hat{c}^T x_{\hat{c}} + \phi(x_{\hat{c}}) \leq t \cdot \hat{c}^T x_c + \phi(x_c)$$

The difference between $f_t(x_{\hat{c}})$ and $f_t(x_c)$ is

$$\begin{aligned} t c^T x_{\hat{c}} + \phi(x_{\hat{c}}) - t c^T x_c - \phi(x_c) \\ \leq t(c^T x_{\hat{c}} + \hat{c}^T x_c - \hat{c}^T x_{\hat{c}} - c^T x_c) \\ \leq 4tn2^{3L} \end{aligned}$$

For $t = 1/2^{\Omega(L)}$ the last term becomes constant. Hence, using damped Newton we can move from $x_{\hat{c}}$ to x_c quickly.

In total for this analysis we require $\mathcal{O}(\sqrt{m}L)$ outer iterations for the whole algorithm.

One iteration can be implemented in $\tilde{O}(m^3)$ time.

How to get close to analytic center?

Clearly,

$$t \cdot \hat{c}^T x_{\hat{c}} + \phi(x_{\hat{c}}) \leq t \cdot \hat{c}^T x_c + \phi(x_c)$$

The difference between $f_t(x_{\hat{c}})$ and $f_t(x_c)$ is

$$\begin{aligned} t c^T x_{\hat{c}} + \phi(x_{\hat{c}}) - t c^T x_c - \phi(x_c) \\ \leq t(c^T x_{\hat{c}} + \hat{c}^T x_c - \hat{c}^T x_{\hat{c}} - c^T x_c) \\ \leq 4tn2^{3L} \end{aligned}$$

For $t = 1/2^{\Omega(L)}$ the last term becomes constant. Hence, using damped Newton we can move from $x_{\hat{c}}$ to x_c quickly.

In total for this analysis we require $\mathcal{O}(\sqrt{m}L)$ outer iterations for the whole algorithm.

One iteration can be implemented in $\tilde{\mathcal{O}}(m^3)$ time.

How to get close to analytic center?

Clearly,

$$t \cdot \hat{c}^T x_{\hat{c}} + \phi(x_{\hat{c}}) \leq t \cdot \hat{c}^T x_c + \phi(x_c)$$

The difference between $f_t(x_{\hat{c}})$ and $f_t(x_c)$ is

$$\begin{aligned} t c^T x_{\hat{c}} + \phi(x_{\hat{c}}) - t c^T x_c - \phi(x_c) \\ \leq t(c^T x_{\hat{c}} + \hat{c}^T x_c - \hat{c}^T x_{\hat{c}} - c^T x_c) \\ \leq 4tn2^{3L} \end{aligned}$$

For $t = 1/2^{\Omega(L)}$ the last term becomes constant. Hence, using damped Newton we can move from $x_{\hat{c}}$ to x_c quickly.

In total for this analysis we require $\mathcal{O}(\sqrt{m}L)$ outer iterations for the whole algorithm.

One iteration can be implemented in $\tilde{\mathcal{O}}(m^3)$ time.

How to get close to analytic center?

Clearly,

$$t \cdot \hat{c}^T x_{\hat{c}} + \phi(x_{\hat{c}}) \leq t \cdot \hat{c}^T x_c + \phi(x_c)$$

The difference between $f_t(x_{\hat{c}})$ and $f_t(x_c)$ is

$$\begin{aligned} t c^T x_{\hat{c}} + \phi(x_{\hat{c}}) - t c^T x_c - \phi(x_c) \\ \leq t(c^T x_{\hat{c}} + \hat{c}^T x_c - \hat{c}^T x_{\hat{c}} - c^T x_c) \\ \leq 4tn2^{3L} \end{aligned}$$

For $t = 1/2^{\Omega(L)}$ the last term becomes constant. Hence, using damped Newton we can move from $x_{\hat{c}}$ to x_c quickly.

In total for this analysis we require $\mathcal{O}(\sqrt{m}L)$ outer iterations for the whole algorithm.

One iteration can be implemented in $\tilde{\mathcal{O}}(m^3)$ time.