Norwegian University of Science and Technology

TTK4135 – Lecture 18
Sequential Quadratic Programming (SQP)

Lecturer: Lars Imsland

Outline

- Recap: Newton's method for solving nonlinear equations
- Recap: Equality-constrained QPs

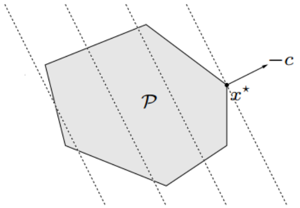
- SQP for *equality-constrained* nonlinear programming problems
  - Next time: SQP for general nonlinear programming problems

Reference: N&W Ch.18-18.1

Types of Constrained Optimization Problems

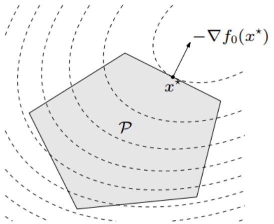
- Linear programming
  - Convex problem
  - Feasible set polyhedron

$$
\begin{array}{rl}
\min & c^\top x \\
\text{subject to} & Ax \le b \\
& Cx = d
\end{array}
$$



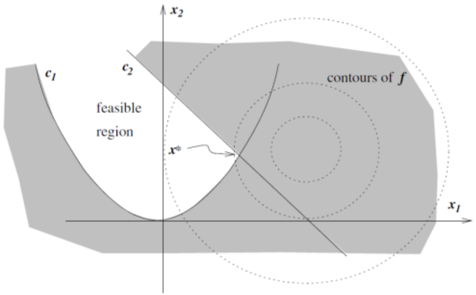
- Quadratic programming
  - Convex problem if $P \geq 0$
  - Feasible set polyhedron

$$
\begin{array}{rl}
\min & \frac{1}{2}x^\top P x + q^\top x \\
\text{subject to} & Ax \le b \\
& Cx = d
\end{array}
$$



- Nonlinear programming
  - In general non-convex!

$$
\begin{array}{rl}
\min & f(x) \\
\text{subject to} & g(x) = 0 \\
& h(x) \ge 0
\end{array}
$$



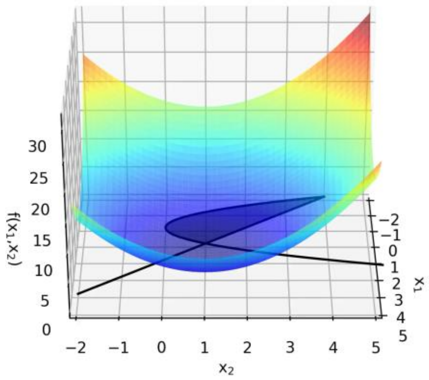
$$
\min_{x \in \mathbb{R}^n} f(x) \quad \text{subject to} \quad \begin{cases} c_i(x) = 0, & i \in \mathcal{E}, \\ c_i(x) \ge 0, & i \in \mathcal{I}. \end{cases}
$$

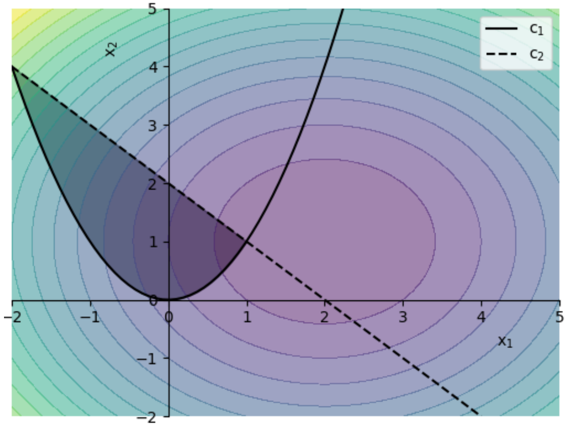
General Optimization Problem (NLP)

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{subject to} \quad \begin{aligned} c_i(x) &= 0, \quad i \in \mathcal{E}, \\ c_i(x) &\geq 0, \quad i \in \mathcal{I}. \end{aligned}$$

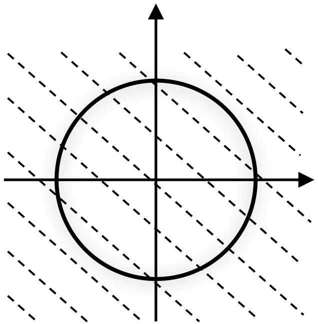
- Example:

$$\min\ (x_1 - 2)^2 + (x_2 - 1)^2 \quad \text{subject to} \quad \begin{aligned} x_1^2 - x_2 &\leq 0, \\ x_1 + x_2 &\leq 2. \end{aligned}$$





Today: Only equality constraints



The Lagrangian

For constrained optimization problems, introduce modification of objective function:

$$\mathcal{L}(x,\lambda) = f(x) - \sum_{i\in\mathcal{E}\cup\mathcal{I}} \lambda_i c_i(x)$$

- Multipliers for *equality* constrains may have both signs in a solution
- Multipliers for *inequality* constraints cannot be negative (cf. shadow prices)
- For (inequality) constraints that are *inactive*, multipliers are zero

KKT conditions (Theorem 12.1)

$$\mathcal{L}(x,\lambda) = f(x) - \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i c_i(x)$$

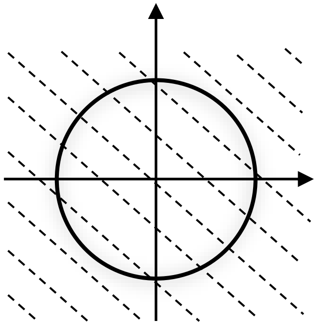
KKT conditions (First-order necessary conditions): If x^* is a local solution and LICQ holds, then there exist λ^* such that

$$\begin{aligned}
\nabla_x \mathcal{L}(x^*, \lambda^*) &= 0, && && \text{(stationarity)} \\
c_i(x^*) &= 0, && \forall i \in \mathcal{E}, && \rbrace\ \text{(primal feasibility)} \\
c_i(x^*) &\geq 0, && \forall i \in \mathcal{I}, && \\
\lambda_i^* &\geq 0, && \forall i \in \mathcal{I}, && \text{(dual feasibility)} \\
\lambda_i^* c_i(x^*) &= 0, && \forall i \in \mathcal{E} \cup \mathcal{I}. && \text{(complementarity condition/ complementary slackness)}
\end{aligned}$$

Either $\lambda_i^* = 0$ *or* $c_i(x^*) = 0$

Example KKT system

$$\min_{x\in\mathbb{R}^2} \; -x_1 - x_2 \quad \text{s.t.} \quad x_1^2 + x_2^2 - 1 = 0$$



Today: Equality-constrained NLP

Newton's method for solving nonlinear equations (Ch. 11)

- Solve equation system $r(x) = 0, \quad r(x) : \mathbb{R}^n \to \mathbb{R}^n$
- Assume Jacobian $J(x) \in \mathbb{R}^{n \times n}$ exists and is continuous
- Taylor: $r(x + p) = r(x) + J(x)p + O(\|p\|^2)$

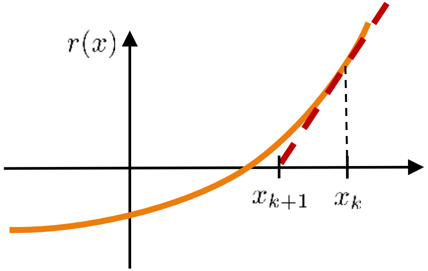
$$J(x) = \begin{pmatrix} \frac{\partial r_1}{\partial x_1} & \frac{\partial r_1}{\partial x_2} & \cdots \\ \frac{\partial r_2}{\partial x_1} & \frac{\partial r_2}{\partial x_2} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}$$

Algorithm 11.1 (Newton's Method for Nonlinear Equations).

Choose x_0;
for $k = 0, 1, 2, \ldots$
Calculate a solution p_k to the Newton equations

$$J(x_k)p_k = -r(x_k);$$

$x_{k+1} \leftarrow x_k + p_k$;
end (for)



- Convergence rate (Thm 11.2): Quadratic convergence if $J(x)$ is invertible
  (quadratic convergence is very good, but only holds close to the solution)

Newton's method to solve $\mathbf{F}(\mathbf{x}, \lambda) = \mathbf{0}$

$$F(x, \lambda) = \begin{pmatrix} \nabla f(x) - A^\top(x)\lambda \\ c(x) \end{pmatrix}$$

Newton's method to solve $\mathbf{F}(\mathbf{x}, \lambda) = \mathbf{0}$

$$F(x, \lambda) = \begin{pmatrix} \nabla f(x) - A^\top(x)\lambda \\ c(x) \end{pmatrix}$$

Equality-constrained QP (EQP)

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & \frac{1}{2} x^\top G x + c^\top x \\ \text{subject to} \quad & Ax = b, \quad A \in \mathbb{R}^{m \times n} \end{aligned}$$

Basic assumption:
A full row rank

- KKT-conditions (KKT system, KKT matrix):

$$\begin{pmatrix} G & -A^\top \\ A & 0 \end{pmatrix} \begin{pmatrix} x^* \\ \lambda^* \end{pmatrix} = \begin{pmatrix} -c \\ b \end{pmatrix} \quad \text{or, if we let } x^* = x + p, \quad \begin{pmatrix} G & A^\top \\ A & 0 \end{pmatrix} \begin{pmatrix} -p \\ \lambda^* \end{pmatrix} = \begin{pmatrix} c + Gx \\ Ax - b \end{pmatrix}$$

- Solvable when $Z^\top G Z > 0$ (columns of *Z* basis for nullspace of *A*)

- That is: QP with only equality constraints is solved by a solving a set of linear equations

Alternative “derivation” of KKT-system

$$\min_{x \in \mathbb{R}^n} \; f(x) \quad \text{s.t.} \quad c(x) = 0$$

Alternative “derivation” of KKT-system, cont’d

From Newton’s method:

$$\underbrace{\begin{pmatrix} \nabla^2_{xx}\mathcal{L}(x_k,\lambda_k) & -A^\top(x_k) \\ A(x_k) & 0 \end{pmatrix}}_{\text{Jacobian of } F(x,\lambda) \text{ at } (x_k,\,\lambda_k)} \begin{pmatrix} p_k \\ p_{\lambda_k} \end{pmatrix} = \underbrace{\begin{pmatrix} -\nabla f(x_k) + A^\top(x_k)\lambda_k \\ -c(x_k) \end{pmatrix}}_{-F(x_k,\lambda_k)}$$

We see that one iteration of algorithm has two interpretations:

1. Newton's method to solve KKT of NLP
   - Analysis: Method has quadratic convergence
2. Sequentially solving QP approximations of NLP
   - Extension to inequalities
   - Practical implementation: Use QP-solvers

Local SQP-algorithm for solving equality-constrained NLPs

$$\min f(x)$$
$$\text{subject to } c(x) = 0$$

Algorithm 18.1 (Local SQP Algorithm for solving (18.1)).

Choose an initial pair (x_0, λ_0); set $k \leftarrow 0$;
repeat until a convergence test is satisfied
  Evaluate f_k, ∇f_k, $\nabla^2_{xx}\mathcal{L}_k$, c_k, and A_k;
  Solve (18.7) to obtain p_k and l_k;
  Set $x_{k+1} \leftarrow x_k + p_k$ and $\lambda_{k+1} \leftarrow l_k$;
end (repeat)

EQP:

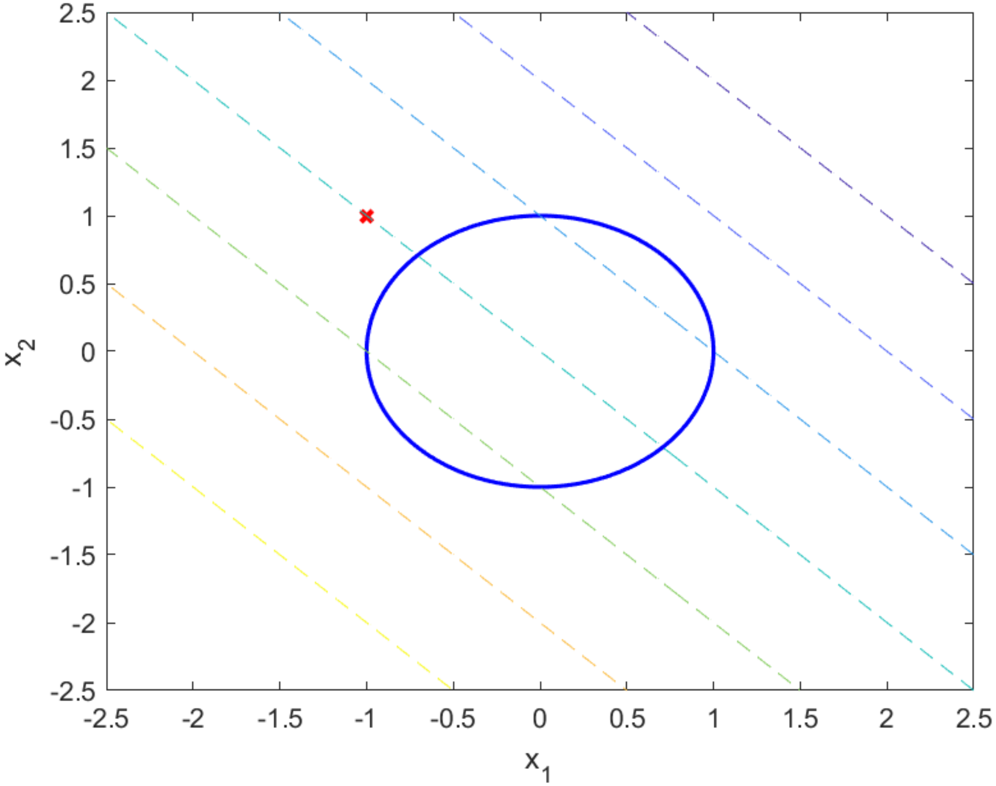
$$\begin{aligned} \min_p \quad & f_k + \nabla f_k^T p + \tfrac{1}{2} p^T \nabla^2_{xx}\mathcal{L}_k p \\ \text{subject to} \quad & A_k p + c_k = 0. \end{aligned}$$

$$\min_{x\in\mathbb{R}^2} \; -x_1 - x_2 \quad \text{s.t.} \quad x_1^2 + x_2^2 - 1 = 0$$

```
% min -x1 - x2 s.t. x1^2 + x2^2 = 1

f = @(x) - x(1) - x(2);
df = @(x) [-1; -1];

c = @(x) x(1)^2 + x(2)^2 - 1;
A = @(x) [2*x(1), 2*x(2)];

HL = @(x,lambda) diag([- 2*lambda, -2*lambda]);

x0 = [-1;1];lambda0 = -1;
x(:,1) = x0; lambda(1,:) = lambda0;

for i = 1:10,
    [p,fval,exitflag,output,lo] = quadprog(HL(x(:,i),lambda(i)),df(x(:,i))',[],[], A(x(:,i)),-c(x(:,i)));
    l = -lo.eqlin;

    % z = [ HL(x(:,i),lambda(i)), -A(x(:,i))'; A(x(:,i)), 0] \ [-df(x(:,i)); -c(x(:,i))];
    % p = z(1:2);
    % l = z(3);

    x(:,i+1) = x(:,i) + p;
    lambda(:,i+1) = l;
end
```

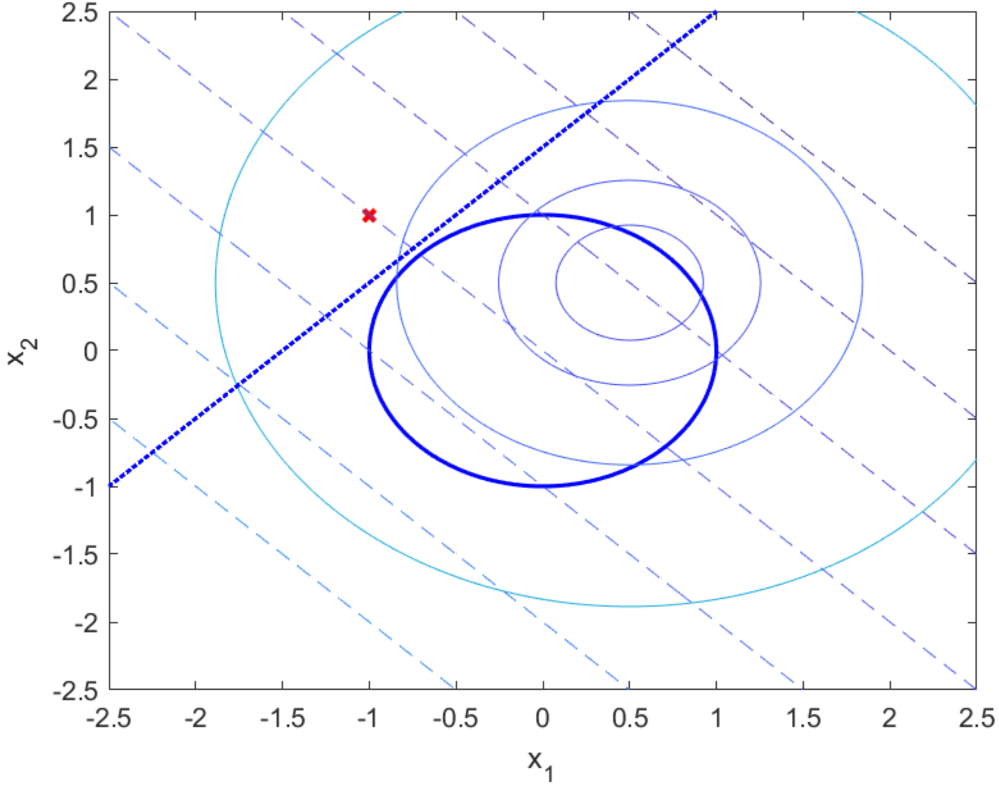
$$\begin{aligned} \min_{p} \quad & f_k + \nabla f_k^T p + \tfrac{1}{2} p^T \nabla_{xx}^2 \mathcal{L}_k p \\ \text{subject to} \quad & A_k p + c_k = 0. \end{aligned}$$

$$\min_{x \in \mathbb{R}^2} \; -x_1 - x_2 \quad \text{s.t.} \quad x_1^2 + x_2^2 - 1 = 0$$



Iteration 1

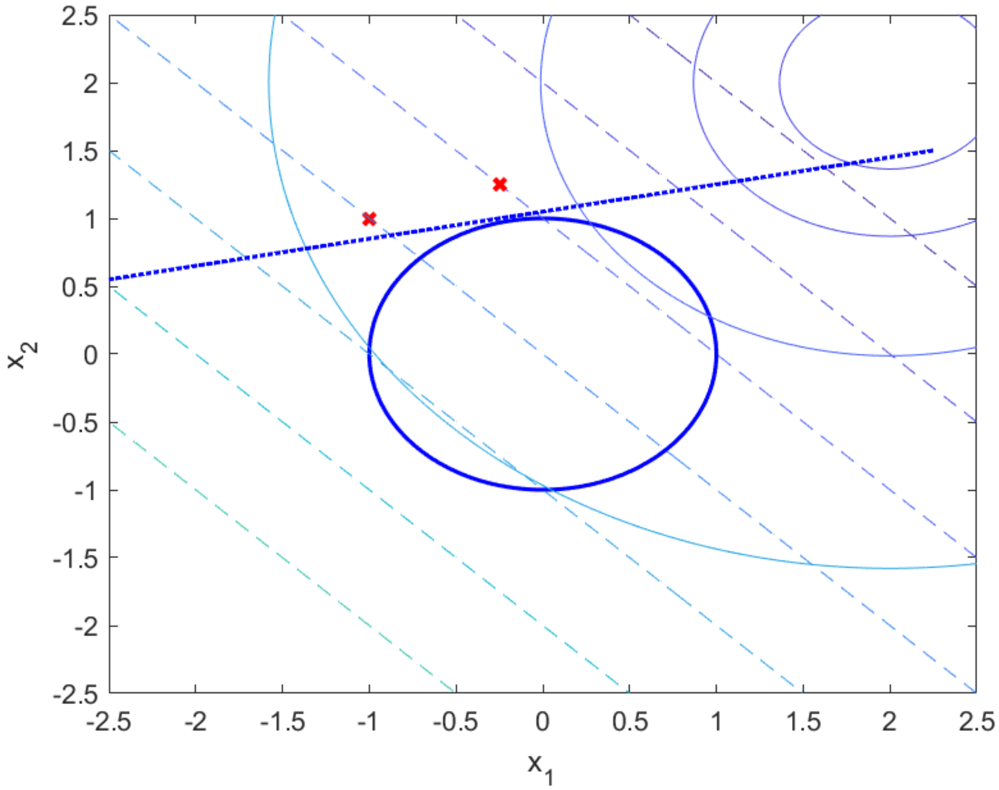
$$\min_{x\in\mathbb{R}^2} \; -x_1 - x_2 \quad \text{s.t.} \quad x_1^2 + x_2^2 - 1 = 0$$



$$\begin{aligned} \min_{p} \quad & f_k + \nabla f_k^T p + \tfrac{1}{2} p^T \nabla_{xx}^2 \mathcal{L}_k p \\ \text{subject to} \quad & A_k p + c_k = 0. \end{aligned}$$

Iteration 1

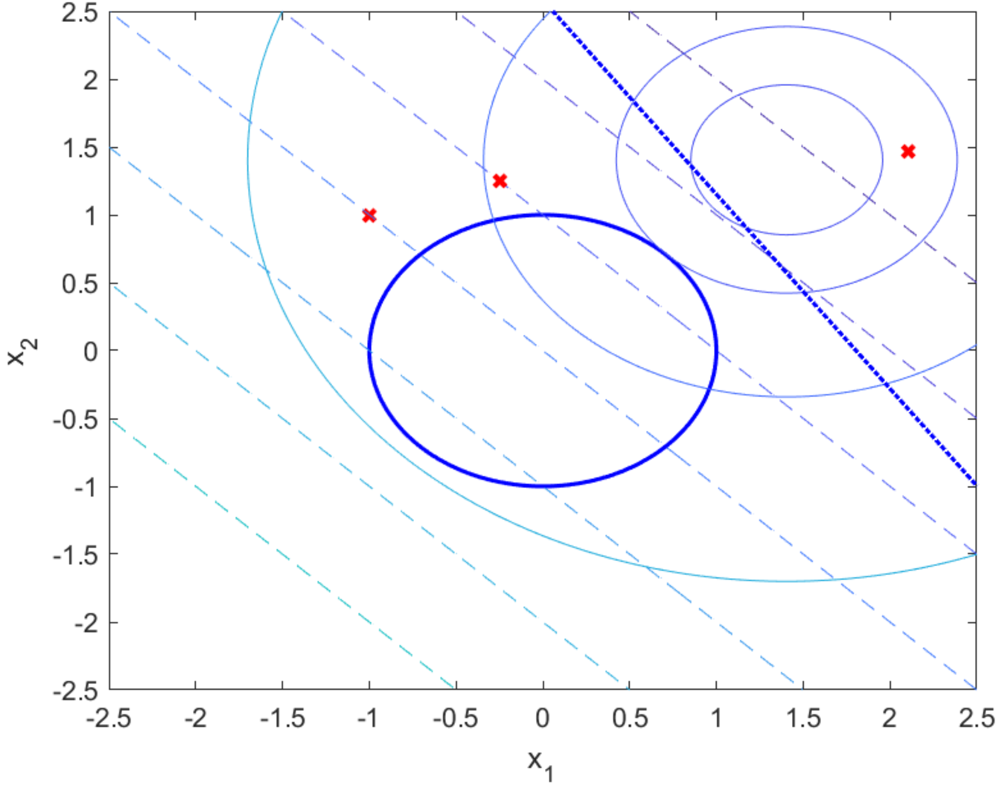
$$\min_{x \in \mathbb{R}^2} \; -x_1 - x_2 \quad \text{s.t.} \quad x_1^2 + x_2^2 - 1 = 0$$



$$\begin{aligned} \min_{p} \quad & f_k + \nabla f_k^T p + \tfrac{1}{2} p^T \nabla_{xx}^2 \mathcal{L}_k p \\ \text{subject to} \quad & A_k p + c_k = 0. \end{aligned}$$

Iteration 3

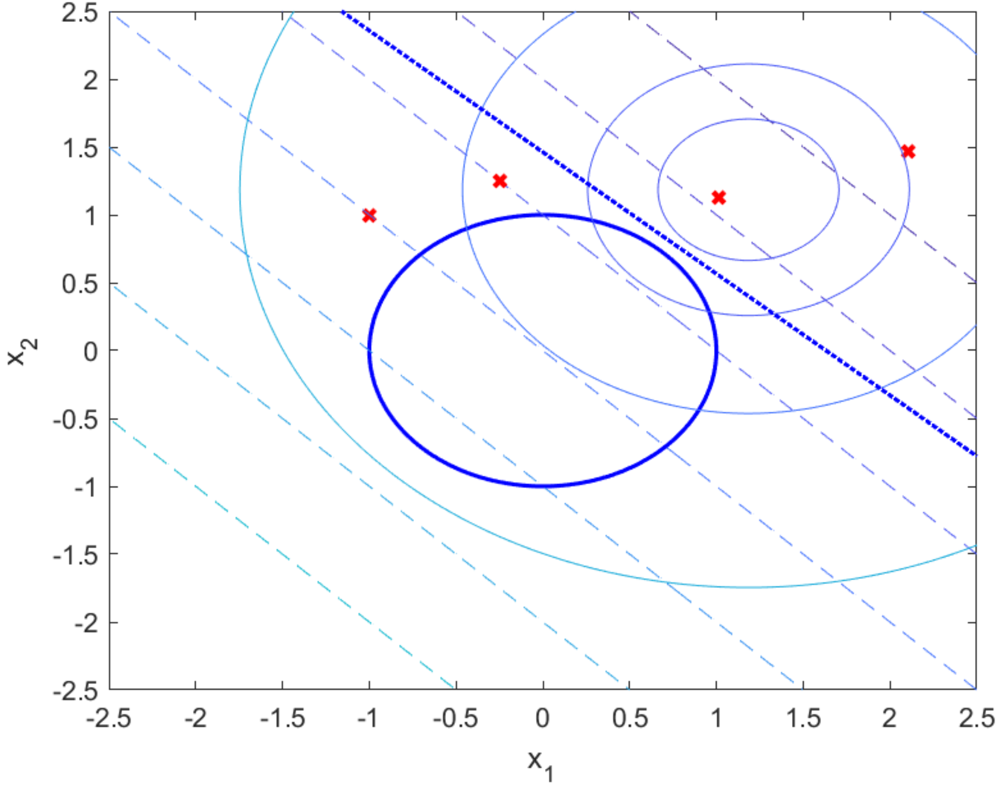
$$\min_{x \in \mathbb{R}^2} \; -x_1 - x_2 \quad \text{s.t.} \quad x_1^2 + x_2^2 - 1 = 0$$



$$\begin{aligned} \min_{p} \quad & f_k + \nabla f_k^T p + \tfrac{1}{2} p^T \nabla_{xx}^2 \mathcal{L}_k p \\ \text{subject to} \quad & A_k p + c_k = 0. \end{aligned}$$

Iteration 4

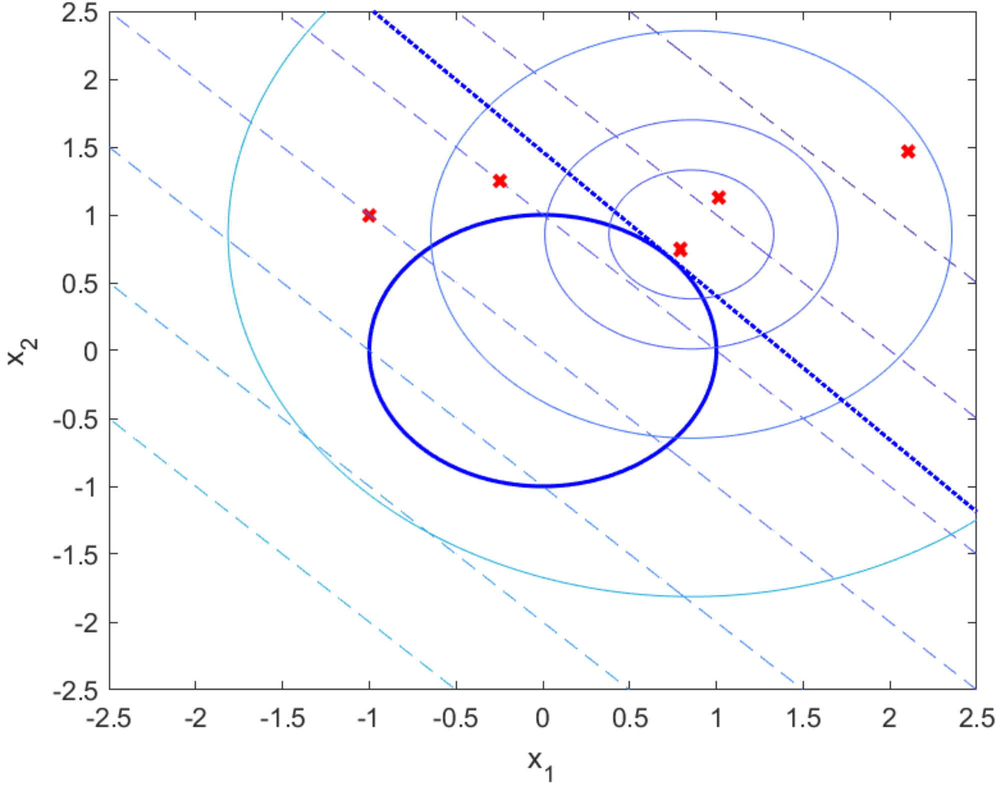
$$\min_{x \in \mathbb{R}^2} \; -x_1 - x_2 \quad \text{s.t.} \quad x_1^2 + x_2^2 - 1 = 0$$



$$\begin{aligned} \min_{p} \quad & f_k + \nabla f_k^T p + \tfrac{1}{2} p^T \nabla_{xx}^2 \mathcal{L}_k p \\ \text{subject to} \quad & A_k p + c_k = 0. \end{aligned}$$

Iteration 5

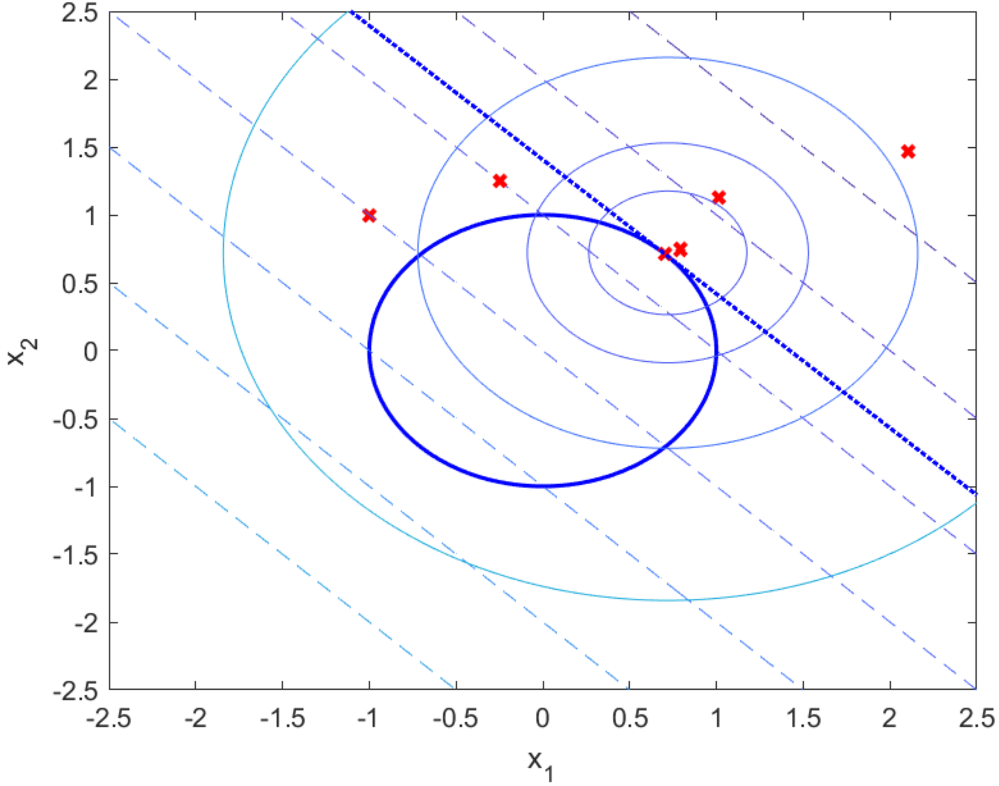
$$\min_{x \in \mathbb{R}^2} \ -x_1 - x_2 \quad \text{s.t.} \quad x_1^2 + x_2^2 - 1 = 0$$



$$\begin{aligned} \min_{p} \quad & f_k + \nabla f_k^T p + \tfrac{1}{2} p^T \nabla_{xx}^2 \mathcal{L}_k p \\ \text{subject to} \quad & A_k p + c_k = 0. \end{aligned}$$

Iteration 6

$$\min_{x \in \mathbb{R}^2} \; -x_1 - x_2 \quad \text{s.t.} \quad x_1^2 + x_2^2 - 1 = 0$$



$$\begin{aligned} \min_{p} \quad & f_k + \nabla f_k^T p + \tfrac{1}{2} p^T \nabla_{xx}^2 \mathcal{L}_k p \\ \text{subject to} \quad & A_k p + c_k = 0. \end{aligned}$$

QP approximation can be seen as approximation of Lagrangian

$$\begin{aligned} \min_{p} \quad & f_k + \nabla f_k^T p + \tfrac{1}{2} p^T \nabla_{xx}^2 \mathcal{L}_k p \\ \text{subject to} \quad & A_k p + c_k = 0. \end{aligned}$$