SI152: Numerical Optimization

Lecture 14: Penalty Methods

Hao Wang

Email: haw309@gmail.com

ShanghaiTech University

November 18, 2025

1 Quadratic Penalization

2 Exact Penalization

Reformulate the constrained problem

$$\min_x f(x) \quad \text{s.t. } c(x) = 0$$

as the unconstrained quadratic penalty subproblem

$$\min_x \phi(x;\nu) = f(x) + \frac{\nu}{2}\|c(x)\|_2^2 = f(x) + \sum_{i\in\mathcal{E}} c_i^2(x)$$

where $\nu \geq 0$ is a penalty parameter.
Recall the example

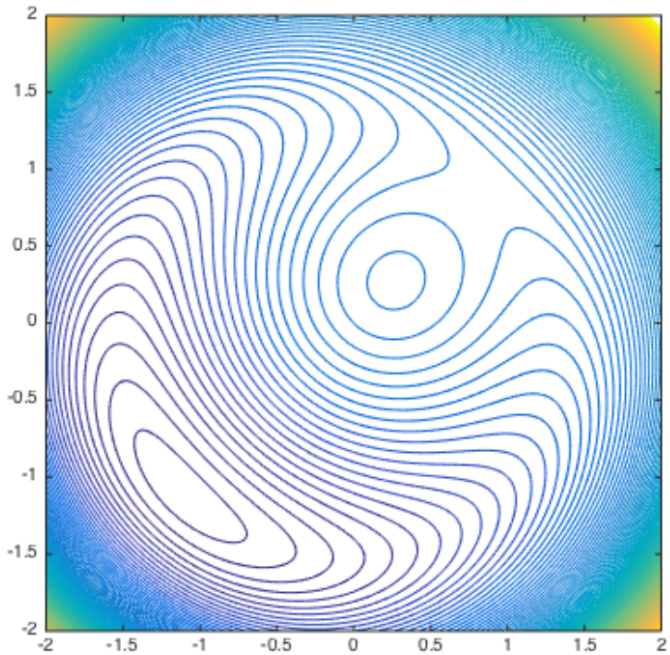
$$\min_x x_1 + x_2 \quad \text{s.t. } x_1^2 + x_2^2 - 2 = 0$$

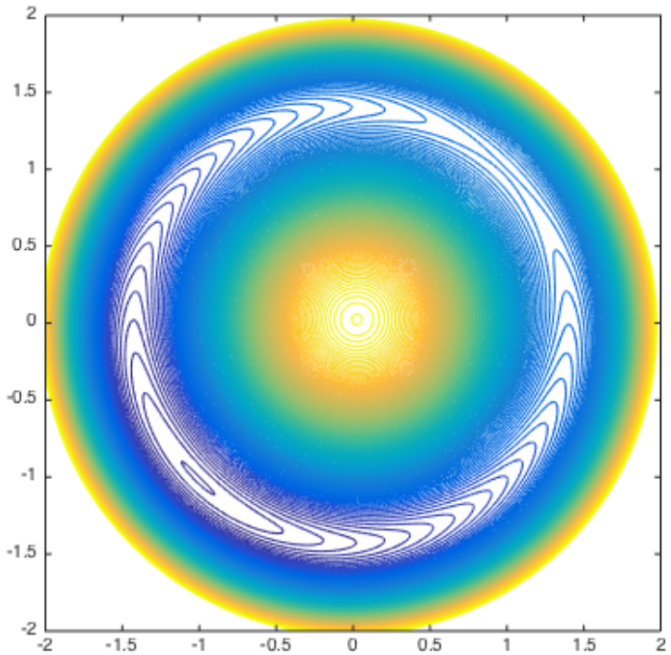
for which we have the corresponding quadratic penalty subproblem

$$\min_x x_1 + x_2 + \frac{\nu}{2}(x_1^2 + x_2^2 - 2)^2$$

- The solution to the constrained problem is $x_* = (-1, -1)$
- The solution to the quadratic penalty subproblem depends on ν

Contours of ϕ for $\nu = 1$ and $\nu = 10$





As $\nu \to \infty$, minimizer of quadratic penalty approaches x_*.

1: Choose $\nu_0 > 0$ and $\{\tau_k\} \to 0$.
2: **for** $k = 0, 1, 2, \ldots$ **do**
3: Find an approximate solution x_k to

$$\min_x \phi(x; \nu_k) = f(x) + \frac{\nu_k}{2} \|c(x)\|_2^2$$

satisfying

$$\|\nabla \phi(x_k; \nu_k)\| \leq \tau_k$$

4: If an optimality test for the constrained problem is satisfied, then stop.
5: Choose $\nu_{k+1} > \nu_k$.
6: **end for**

Theorem 1

Suppose $\nu_k \to \infty$ and each x_k is the exact global minimizer of $\phi(\cdot; \nu_k)$. Then every limit point of $\{x_k\}$ is a global solution of the constrained problem.

- Nice! But...
- Unfortunately, finding exact global minimizers for each ν_k is expensive.
- There are other concerns, too, but first a more practical result...

Definition 2 (Infeasible Stationary Point)

If a point is infeasible for the constrained problem, but is stationary for the infeasibility measure $\|c(x)\|_2^2$, then it is an infeasible stationary point.

Theorem 3

Suppose $\nu_k \to \infty$ and $\tau_k \to 0$. Let x_* be any limit point of $\{x_k\}$.

- *If x_* is infeasible, then it is a stationary point of $\|c(x)\|^2$.*
- *If x_* is is feasible and the constraint gradients are linearly independent, then x_* is a KKT point for the constrained problem. Moreover, for such points, we have that for any infinite subsequence $\mathcal{K}$ such that*

$$\lim_{k \in \mathcal{K}} x_k = x_*$$

the following limit holds

$$\lim_{k \in \mathcal{K}} \nu_k c_i(x_k) = \lambda_i^*,$$

where λ^* is a multiplier vector satisfying the KKT condition

$$\nabla f(x_*) + \nabla c(x_*)\lambda_* = 0.$$

First some comments, then a quick proof.

- We cannot guarantee that we find a feasible point.
- We only guarantee that we find a stationary point of $\|c(x)\|^2$, i.e., a point with

$$\nabla c(x) c(x) = \sum c_i(x) \nabla c_i(x) = 0.$$

- This may mean $c_i(x) = 0$ for $i \in \mathcal{E}$, but perhaps not if the constraint gradients are linearly dependent for some x.
- If we converge to a feasible point, then we need the LICQ to prove it's a KKT point. The multipliers are then revealed by the penalty parameter and constraint values.

Proof.

- By design of the algorithm, x^k satisfies

$$\|\nabla\phi(x^k;\nu_k)\| = \|\nabla f(x^k) + \sum \nu^k c_i(x^k)\nabla c_i(x^k)\| \leq \tau_k.$$

- Rearranging, and using the inequality $\|a\| - \|b\| \leq \|a+b\|$, we find

$$\|\sum c_i(x^k)\nabla c_i(x^k)\| \leq \frac{1}{\nu_k}(\tau_k + \|\nabla f(x^k)\|).$$

- Since $\nu_k \to \infty$ and $\nabla f(x^k) \to \nabla f(x^*)$ for $k \in \mathcal{K}$, this implies

$$\|\sum c_i(x^*)\nabla c_i(x^*)\| = 0.$$

- This proves the infeasible case, and proves that we converge to a feasible point if the constraint gradients are linearly independent.

□

Proof.

It remains to show that under LICQ we obtain a KKT point with the given λ^*

- Define $\lambda^k = \nu_k c(x^k)$, which by definition of ϕ means that for all k

$$\nabla c(x^k)\lambda^k = -(\nabla f(x^k) - \nabla \phi(x^k; \nu_k)).$$

- For $k \in \mathcal{K}$ sufficiently large, $\nabla c(x^k)$ has full rank, so we can write

$$\lambda^k = -[\nabla c(x^k)^T \nabla c(x^k)]^{-1} \nabla c(x^k)^T (\nabla f(x^k) - \nabla \phi(x^k; \nu_k)).$$

- Taking limits on both sides, we obtain

$$\lim_{k \in \mathcal{K}} \lambda^k = -[\nabla c(x^*)^T \nabla c(x^*)]^{-1} \nabla c(x^*)^T \nabla f(x^*) =: \lambda^*$$

  which implies that

$$\nabla f(x^*) + \nabla c(x^*)\lambda^* = 0.$$

Corresponding to the generally constrained problem

$$\begin{aligned} \min_{x} \quad & f(x) \\ \text{s.t.} \quad & c_{\mathcal{E}}(x) = 0 \\ & c_{\mathcal{I}}(x) \leq 0, \end{aligned}$$

we have the following quadratic penalty subproblem

$$\begin{aligned} \min_{x} \phi(x;\nu) :=& f(x) + \frac{\nu}{2}\|c_{\mathcal{E}}(x)\|_2^2 + \frac{\nu}{2}\|\max\{c_{\mathcal{I}}(x), 0\}\|_2^2 \\ =& f(x) + \frac{\nu}{2}\sum_{i\in\mathcal{E}} c_i^2(x) + \frac{\nu}{2}\sum_{i\in\mathcal{E}} \max\{c_i(x), 0\}^2. \end{aligned}$$

Results similar to those discussed hold in the generally constrained case as well.

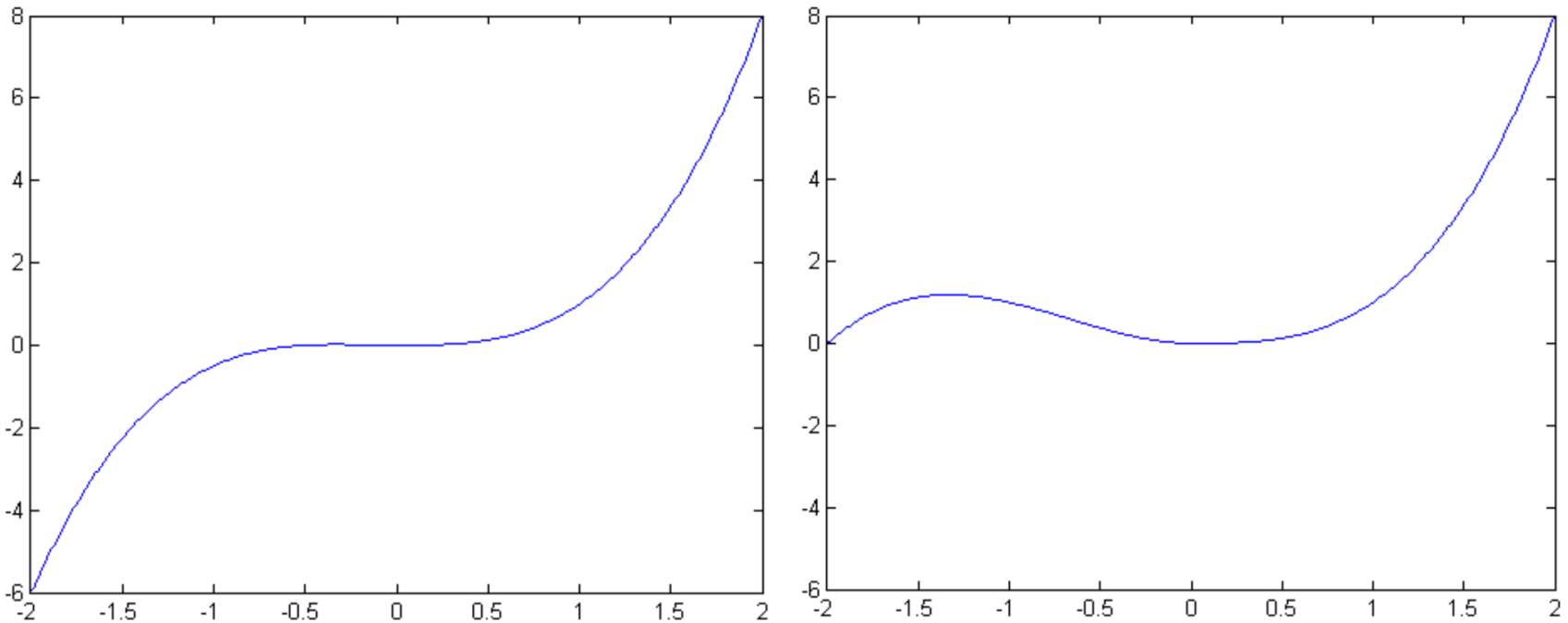
Despite some nice features, there are major drawbacks to quadratic penalization.

- How do we define $\{\nu_k\}$ so that we converge to a feasible point?
- We may have a hard time finding x^k satisfying

$$\|\nabla\phi(x^k;\nu_k)\| \leq \tau_k.$$

- Quadratic penalization leads to terrible ill-conditioning.

$\min_x x^3$ s.t. $-x \leq 0$ yields $\phi(x;\nu) = x^3 + \frac{\nu}{2}\max\{-x,0\}^2$. (Here, $\nu = 1, 4$):

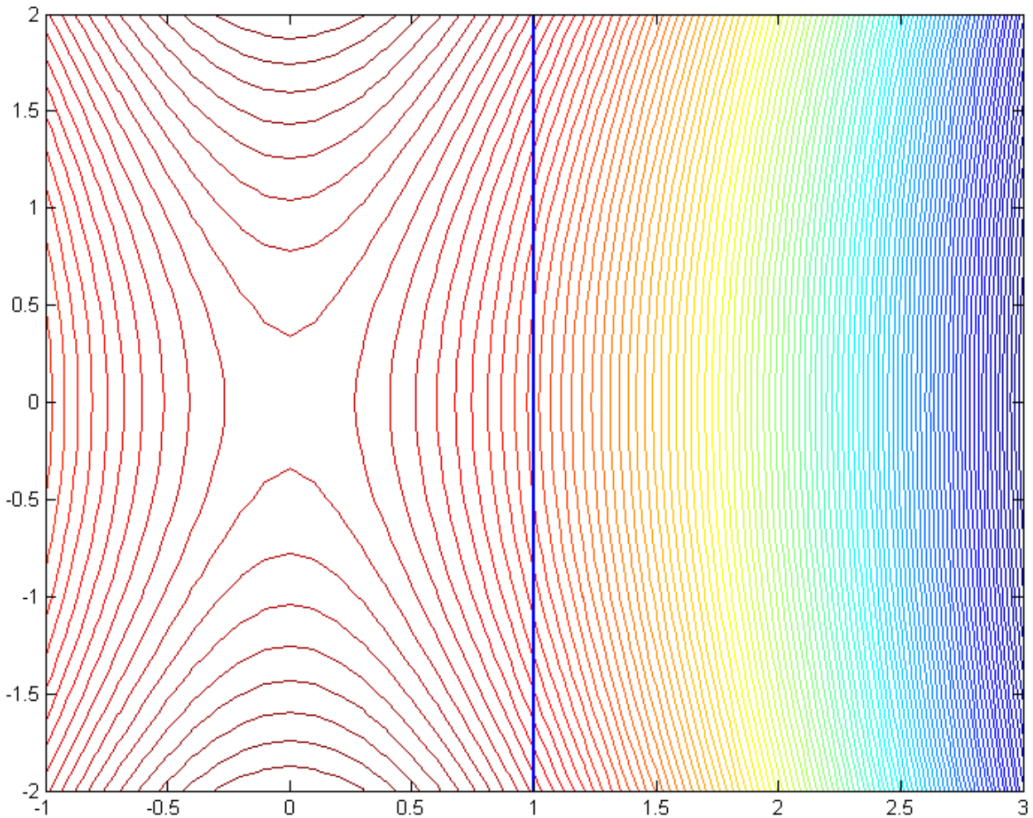


f is convex on the feasible set, but ϕ is unbounded below for all ν!

The following constrained problem (convex on the feasible set!) has $x^* = (1, 0)$.

$$\min_x \; f(x) = -5x_1^2 + x_2^2$$

$$\text{s.t. } x_1 - 1 = 0.$$

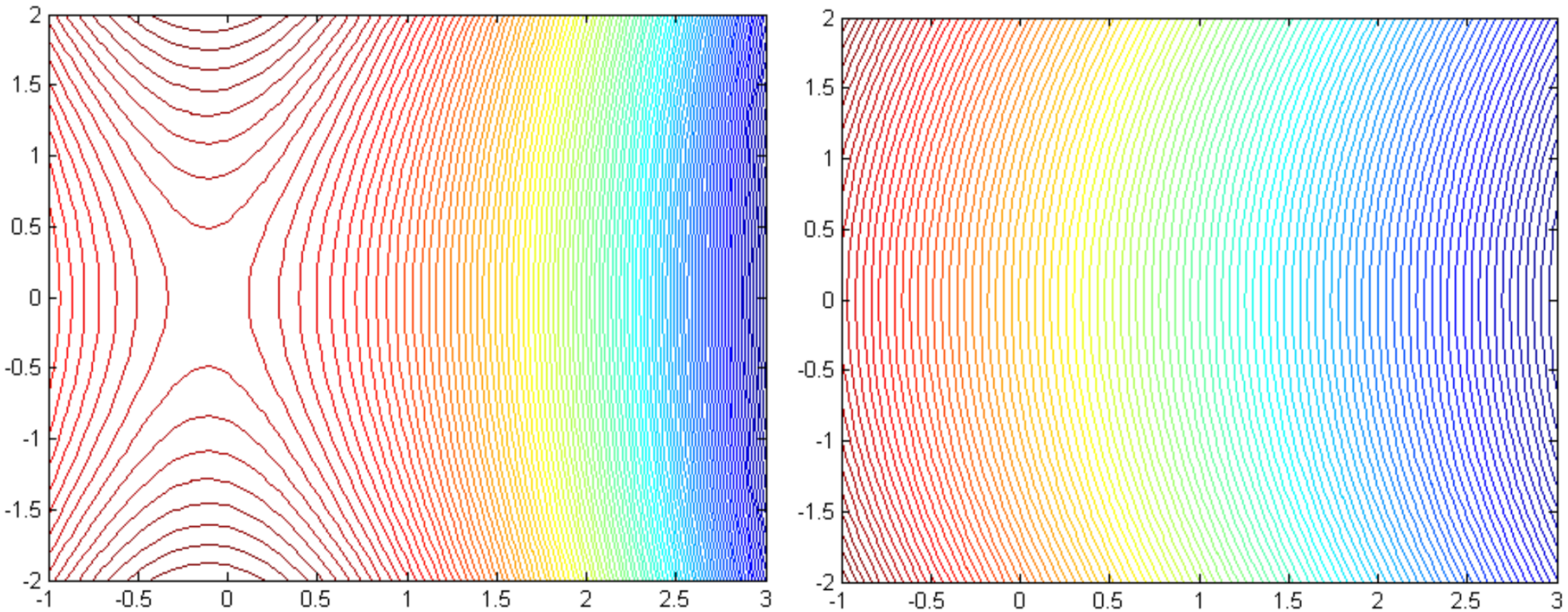


f is convex on the feasible set, but ϕ is unbounded below for all $\nu \in (0, 10)$!

However, the corresponding quadratic penalty subproblem

$$\min_x \phi(x;\nu) = -5x_1^2 + x_2^2 + \frac{\nu}{2}(x_1 - 1)^2$$

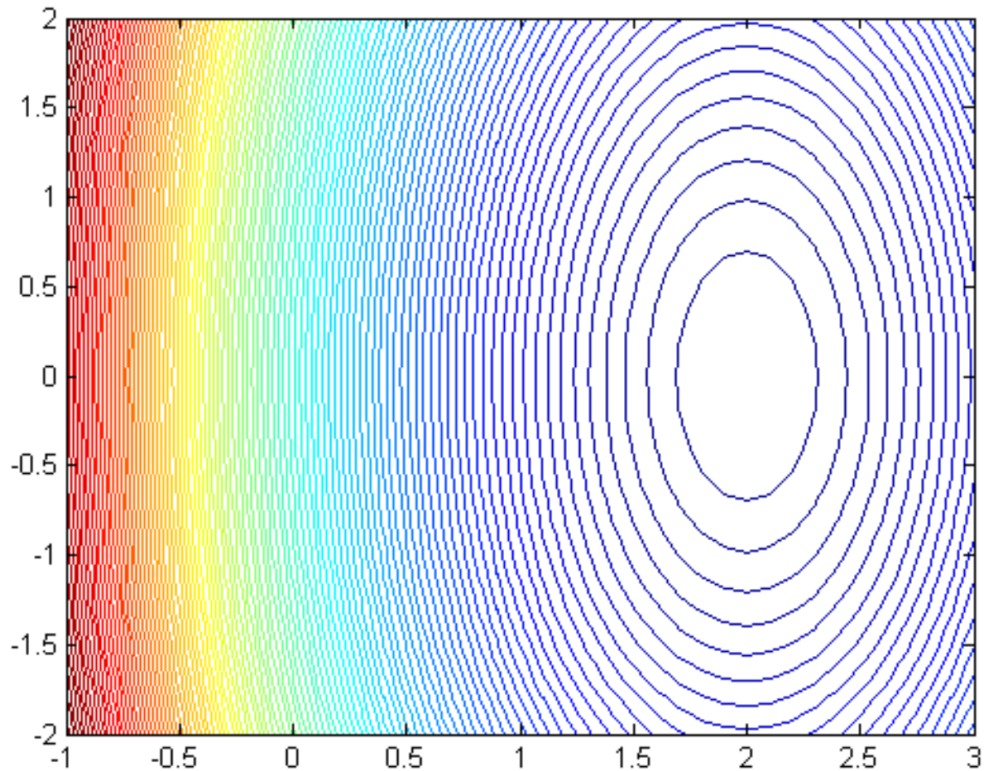
is unbounded below for all $\nu < 10$. (Below, $\nu = 1, 10$.)

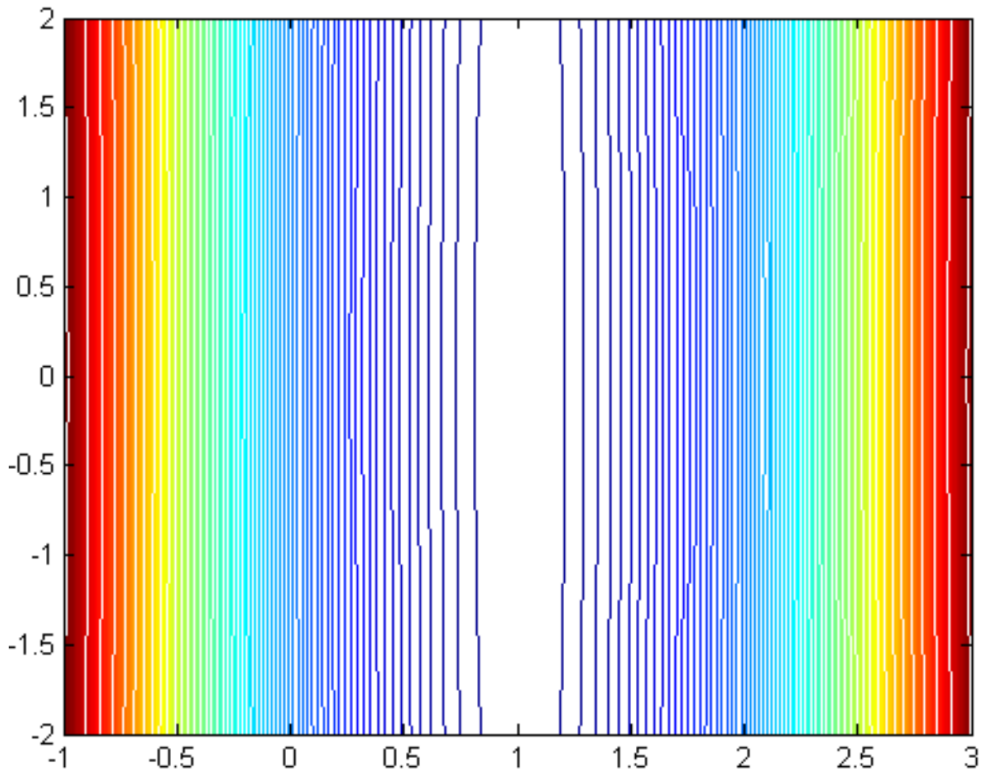


However, the corresponding quadratic penalty subproblem

$$\min_x \phi(x;\nu) = -5x_1^2 + x_2^2 + \frac{\nu}{2}(x_1 - 1)^2$$

is unbounded below for all $\nu < 10$. (Below, $\nu = 20, 1000$.)





It is apparent in this last example that the Hessian of the quadratic penalty function can become extremely ill-conditioned as $\nu \to \infty$:

$$\begin{aligned}\nabla^2 \phi(x;\nu) &= \nabla^2 f(x) + \sum \nu c_i(x) \nabla c_i(x) + \nu \nabla c_i(x) \nabla c_i(x)^T \\ &= \nabla^2_{xx} L(x,\lambda) + \nu \sum \nabla c_i(x) \nabla c_i(x)^T.\end{aligned}$$

- Some eigenvalues of $\nabla^2 \phi(x;\nu)$ increase with ν, even if f and c are "nice".
- The multiplier estimate $\lambda^k = \nu_k c(x^k)$ may be a poor approximation for the optimal multipliers, which may adversely affect convergence of an unconstrained optimization method applied to the quadratic penalty subproblem.

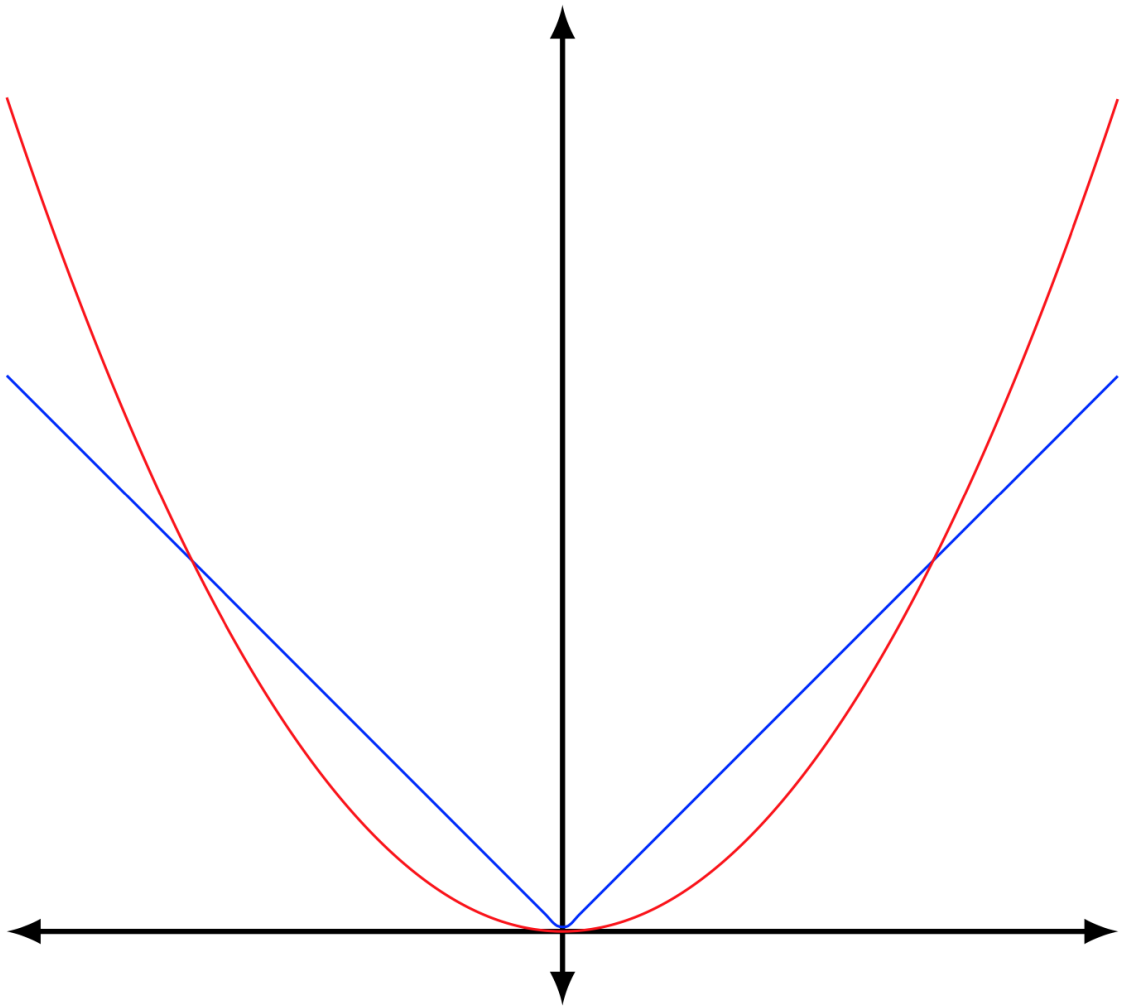
1 Quadratic Penalization

2 Exact Penalization

- The quadratic penalty function

$$\phi(x;\nu) = f(x) + \frac{\nu}{2}\|c(x)\|_2^2 = \sum_{i\in\mathcal{E}} c_i^2(x)$$

  places a large penalty on points with large violations in the constraints.
- However, it is also true that small violations are not penalized very much.
- Thus, we are required to have $\nu \to \infty$ in order to truly satisfy the constraints.
- We will find a better trade-off by penalizing large violations less while penalizing small violations more than a quadratic penalty function.
- This will lead to better behavior for finite ν.



Definition 4 (Exact Penalty Function)

A penalty function $\phi(x;\nu)$ is exact if there exists ν_* such that for all $\nu > \nu_*$, a local solution of the constrained problem is a local minimizer of $\phi(x;\nu)$.

- The quadratic penalty function is not exact.
- The following ℓ_1 penalty function is exact:

$$\phi(x;\nu) := f(x) + \nu\|c(x)\|_1.$$

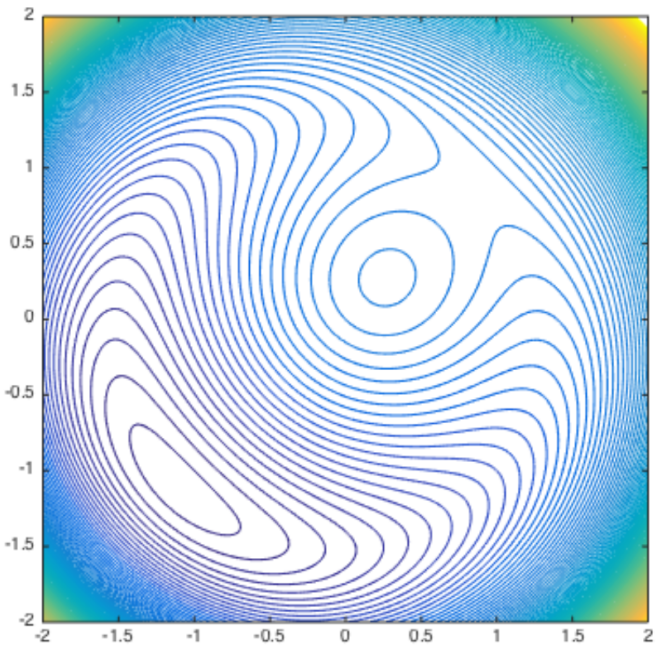
Theorem 5

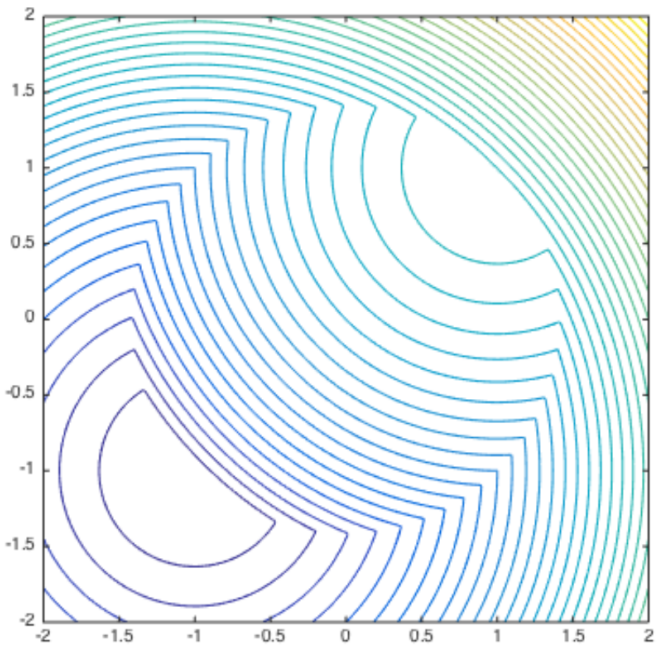
Suppose that x^* is a strict local solution of the constrained problem at which the KKT conditions are satisfied with Lagrange multipliers λ^*. Then, x^* is a local minimizer of $\phi(x;\nu)$ for all $\nu > \nu_*$ where

$$\nu_* = \|\lambda^*\|_\infty.$$

If, in addition, the second-order sufficient conditions holds at x^* and $\nu > \nu_*$, then x^* is a strict local minimizer of $\phi(x;\nu)$.

Contours of quadratic penalty (left) and ℓ_1 penalty (right) for $\nu = 1$.





ℓ_1 penalizes small violations more than quadratic penalization.

We can now use a similar strategy as before, except that if things are nice, then for a finite value of ν we'll obtain the true solution to the constrained problem.

- However, there is a catch... $\phi(x;\nu)$ is nonsmooth!

Theorem 6

A point x^* is stationary for the ℓ_1 penalty function for $\nu = \nu_*$ if the directional derivative of $\phi(x;\nu_*)$ at $x = x_*$ is nonnegative for any d, i.e.,

$$D\phi(d;\nu_*,x^*) = \nabla f(x^*)^T d + \nu_* \sum |\nabla c_i(x^*)^T d| \geq 0, \ \forall d.$$

(Recalling our subdifferential calculus, we know that $\phi(x;\nu)$ admits a directional derivative for any d. Also, note that this definition for $\nu = 0$ provides a definition for a stationary point for the measure of infeasibility $\|c(x)\|_1$.)

Definition 7 (Infeasible Stationary Point)

If a point is infeasible for the constrained problem, but is stationary for the infeasibility measure $\|c(x)\|_1$, then it is an infeasible stationary point.

1: Choose $\nu_0 > 0$ and $\epsilon > 0$.
2: **for** $k = 0, 1, 2, \ldots$ **do**
3: Find an approximate solution x_k to

$$\min_x \phi(x; \nu_k) = f(x) + \nu_k \|c(x)\|_1$$

4: If $\|c(x^k)\|_1 \leq \epsilon$, then stop.
5: Choose $\nu_{k+1} > \nu_k$.
6: **end for**

Theorem 8

Suppose x^k is a stationary point for $\phi(x;\nu)$ for all ν greater than a certain threshold. If x^k is feasible, then there exists λ^k such that (x^k, λ^k) is a KKT point; otherwise, x^k is an infeasible stationary point.

That is, suppose you compute x^k by minimizing $\phi(x;\nu)$ for some "large" ν. If the minimizer doesn't change no matter how much you increase ν, then the result of the above theorem applies.

Do we have the same drawbacks as quadratic penalization?

- We still need to define $\{\nu_k\}$, but the choice is less important as we only need to eventually choose ν_k above a finite threshold.
- However, we may still have a hard time finding x^k that approximately minimizes the penalty function. For example, we may still have $\phi(x;\nu)$ unbounded below, even if f is convex and bounded below on the feasible set.
- Is ill-conditioning a problem? Not as much, again since we only need to eventually choose ν_k above a finite threshold.

Corresponding to the generally constrained problem

$$\begin{aligned} \min_{x} \quad & f(x) \\ \text{s.t.} \quad & c_{\mathcal{E}}(x) = 0 \\ & c_{\mathcal{I}}(x) \leq 0, \end{aligned}$$

we have the following ℓ_1 exact penalty subproblem

$$\begin{aligned} \min_{x} \phi(x;\nu) :=& f(x) + \nu \|c_{\mathcal{E}}(x)\|_1 + \nu \| \max\{c_{\mathcal{I}}(x), 0\}\|_1 \\ =& f(x) + \nu \sum_{i \in \mathcal{E}} |c_i(x)| + \nu \sum_{i \in \mathcal{I}} \max\{c_{\mathcal{I}}(x), 0\}. \end{aligned}$$

Results similar to those discussed hold in the generally constrained case as well.

Penalty function

$$\phi(x;\nu) = f(x) + \nu \sum_{i\in\mathcal{E}} |c_i(x)| + \nu \sum_{i\in\mathcal{I}} \max\{c_{\mathcal{I}}(x), 0\}.$$

Consider the "linearized" penalty function

$$l_k(d) := \nabla f(x^k)^T d + \nu \sum_{i\in\mathcal{E}} |c_i(x^k) + \nabla c_i(x^k)^T d| + \nu \sum_{i\in\mathcal{I}} \max\{c_i(x^k) + \nabla c_i(x^k)^T d, 0\}$$

$$l_k(0) = \nu \sum_{i\in\mathcal{E}} |c_i(x^k)| + \nu \sum_{i\in\mathcal{I}} \max\{c_i(x^k), 0\}$$

$$\begin{aligned} l_k(0) - l_k(d) = & -\nabla f(x^k)^T d \\ & + \nu \sum_{i\in\mathcal{E}} |c_i(x^k) + \nabla c_i(x^k)^T d| + \nu \sum_{i\in\mathcal{I}} \max\{c_i(x^k) + \nabla c_i(x^k)^T d, 0\} \\ & - \nu \sum_{i\in\mathcal{E}} |c_i(x^k)| + \nu \sum_{i\in\mathcal{I}} \max\{c_i(x^k), 0\} \end{aligned}$$

$$D(d; x^k, \nu_k) \le -[l_k(0) - l_k(d)]$$

$$D(d; x^k, \nu_k) = \lim_{\epsilon \to 0} \frac{\phi(x^k + \epsilon d) - \phi(x^k)}{\epsilon}$$

$$\lim_{\epsilon \to 0} \frac{f(x^k + \epsilon d) - f(x^k)}{\epsilon} = \nabla f(x^k)^T d$$

$$\lim_{\epsilon \to 0} \sum_{i \in \mathcal{E}} \frac{|c_i(x^k + \epsilon d)| - |c_i(x^k)|}{\epsilon} = \sum_{i \in \mathcal{E}, c_i^k = 0} |\nabla c_i(x^k)^T d| + \sum_{i \in \mathcal{E}, c_i^k > 0} \nabla c_i(x^k)^T d - \sum_{i \in \mathcal{E}, c_i^k < 0} \nabla c_i(x^k)^T d$$

$$\lim_{\epsilon \to 0} \sum_{i \in \mathcal{E}} \frac{\max\{c_i(x^k + \epsilon d), 0\} - \max\{c_i(x^k), 0\}}{\epsilon} = \sum_{i \in \mathcal{I}, c_i^k = 0} \max\{\nabla c_i(x^k)^T d, 0\} + \sum_{i \in \mathcal{I}, c_i^k > 0} \nabla c_i(x^k)^T d$$

$$\begin{cases} |c_i(x^k) + \nabla c_i(x^k)^T d| - |c_i(x^k)| = |\nabla c_i(x^k)^T d| & \text{if } i \in \mathcal{E}, c_i^k = 0 \\ |c_i(x^k) + \nabla c_i(x^k)^T d| - |c_i(x^k)| \geq \nabla c_i(x^k)^T d & \text{if } i \in \mathcal{E}, c_i^k > 0 \\ |c_i(x^k) + \nabla c_i(x^k)^T d| - |c_i(x^k)| \geq -\nabla c_i(x^k)^T d & \text{if } i \in \mathcal{E}, c_i^k < 0 \end{cases}$$

$$\begin{cases} \max\{c_i(x^k) + \nabla c_i(x^k)^T d, 0\} - \max\{c_i(x^k), 0\} \geq \max\{\nabla c_i(x^k)^T d, 0\} & \text{if } i \in \mathcal{I}, c_i^k = 0 \\ \max\{c_i(x^k) + \nabla c_i(x^k)^T d, 0\} - \max\{c_i(x^k), 0\} \geq \nabla c_i(x^k)^T d & \text{if } i \in \mathcal{I}, c_i^k > 0 \\ \max\{c_i(x^k) + \nabla c_i(x^k)^T d, 0\} - \max\{c_i(x^k), 0\} \geq 0 & \text{if } i \in \mathcal{I}, c_i^k < 0 \end{cases}$$

$$D(d; x^k, \nu_k) \leq \nabla f(x^k)^T d$$

$$+ \sum_{i \in \mathcal{E}} |c_i(x^k) + \nabla c_i(x^k)^T d| + \sum_{i \in \mathcal{I}} \max\{c_i(x^k) + \nabla c_i(x^k)^T d, 0\} - \sum_{i \in \mathcal{E}} |c_i(x^k)| + \sum_{i \in \mathcal{I}} \max\{c_i(x^k), 0\}$$

Therefore, we can minimize the "linearized penalty" to obtain a descent direction for ϕ:

$$\min_d l_k(d) = \nabla f(x^k)^T d + \nu \sum_{i \in \mathcal{E}} |c_i(x^k) + \nabla c_i(x^k)^T d| + \nu \sum_{i \in \mathcal{I}} \max\{c_i(x^k) + \nabla c_i(x^k)^T d, 0\}$$

- This is a linear programming. May need a "trust region" $\|d\| \leq R$ to prevent super long d.
- As long as we have d_k cause a reduction in $l_k(d)$, it is a descent direction for ϕ

$$D(d; x^k, \nu_k) \leq -[l_k(0) - l_k(d)]$$

- This is called a "penalty-SLP (successive linear programming)" algorithm.

Therefore, we can minimize the "linearized penalty" with a quadratic term $H \succ 0$

$$\min_d q_k(d) = \frac{1}{2} d^T H d + l_k(d)$$

- This is a quadratic programming. d can't be too long.
- $q_k(0) = l_k(0)$. As long as we have d_k cause a reduction in $q_k(d)$, we have

  $$0 < q_k(0) - q_k(d) = l_k(0) - l_k(d) - \frac{1}{2} d^T H d \implies l_k(0) - l_k(d) \geq \frac{1}{2} d^T H d > 0.$$

  Therefore, it is a descent direction for ϕ

  $$D(d; x^k, \nu_k) \leq -[l_k(0) - l_k(d)]$$

- This is called a "penalty-SQP (sequential quadratic programming)" algorithm.