Constrained Optimization Problem and Duality

Goals of Lagrange Duality

- Find a lower bound on a minimization problem
- Derive optimality conditions for convex problems
- Get certificate for optimality of a problem
- Remove constraints
- Reformulate problem

Lagrangian and Constrained Optimization I

- Start with a constraint optimization problem

$$\begin{aligned} \min \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \le 0, \quad i = 1, 2, \ldots, m. \end{aligned}$$

- An obvious approach is to formulate the optimization in one function

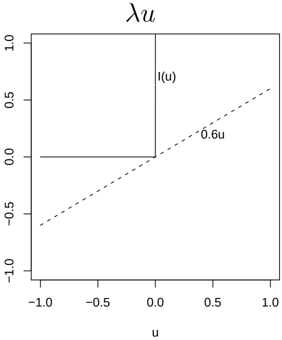
$$\begin{aligned} J(x) &= \begin{cases} f_0(x)\,, & \text{if } f_i(x) \le 0, \forall i \\ \infty \quad , & \text{otherwise.} \end{cases} \\ &= f_0(x) + \sum_{i=1}^{m} I_i[f_i(x)] \end{aligned}$$

where $I[u] = \begin{cases} 0, & \text{if } u \le 0 \\ \infty, & \text{otherwise} \end{cases}$.

- Minimise $J(x)$ to find the solution

Lagrangian and Constrained Optimization II

- However, the indicator function $I[u]$ is non-differentiable and discontinuous
  Hence, hard to optimise $J(x)$
- Replace $I[u]$ by its approximation



For $\lambda > 0$, λu is a lower bound on $I[u]$, a linear relaxation.

Lagrangian and Constrained Optimization III

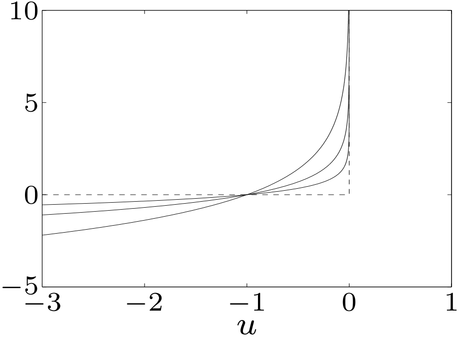
- **Lagrangian**
$$L(x, \lambda) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x)$$
- Note: for a feasible x, i.e., $f_i(x) \leq 0 \ \forall i$, then $L(x, 0) = f_0(x)$
- If $f_i(x) > 0$ for some i, then by taking $\lambda_i \to \infty$, $L(x, \lambda)$ infinite. So
$$\max_{\lambda} L(x, \lambda) = J(x)$$
That means
$$\min_{x} J(x) = \min_{x} \max_{\lambda} L(x, \lambda)$$
- Change the order of maximization over λ and minimization over x. Then
$$\max_{\lambda} \min_{x} L(x, \lambda) = \max_{\lambda} g(\lambda)$$
- **Dual function** $g(\lambda) = \min_x L(x, \lambda)$ is concave since it is the pointwise infimum of Lagrangian which is affine wrt λ
- Maximizing $g(\lambda)$ is a convex optimization problem.

Approximation via Logarithmic barrier

Reformulate the indicator function

$$\min \quad f_0(x) - \frac{1}{t}\sum_i \log(-f_i(x))$$

for $t > 0$, $\frac{-1}{t}\log(-u)$ is a smooth approximation of $I(u)$



Lagrangian

- **Standard optimization problem**: (not necessarily convex)

$$\begin{aligned} \text{minimize}_x \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \le 0,\ i = 1, \ldots, m \\ & h_i(x) = 0,\ i = 1 \ldots, p \end{aligned}$$

  variable $x \in \mathbb{R}^n$, domain $\mathcal{D}$, optimal value $p^\star$

- **Lagrangian** $L : \mathbb{R}^n \times \mathbb{R}^m \times R^p \to R$

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x)$$

  with dom$L = \mathcal{D} \times \mathbb{R}^m \times \mathbb{R}^p$

  - λ_i Lagrangian multiplier associated with the ith inequality $f_i(x) \le 0$
  - ν_i Lagrangian multiplier associated with the ith equality $h_i(x) = 0$
  - λ, ν: Lagrangian multipliers or dual variables

Lagrangian dual function

Lagrangian dual function: $g : \mathbb{R}^m \times \mathbb{R}^p \to R$,

$$
\begin{aligned}
g(\lambda, \nu) &= \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) \\
&= \inf_{x \in \mathcal{D}} f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x)
\end{aligned}
$$

g is concave, can be $-\infty$ for some λ, ν

Lagrangian dual function

Theorem (Weak Duality - lower bounds on optimal value)
For any $\lambda > 0$ and any ν, we have

$$g(\lambda, \nu) \leq f_0(x^\star)$$

Proof. Suppose $\tilde{x}$ is a feasible point. Since $\lambda > 0$,

$$\sum_{i=1}^{m} \lambda_i f_i(\tilde{x}) + \sum_{i=1}^{p} \nu_i h_i(\tilde{x}) \leq 0$$

Hence $L(\tilde{x}, \lambda, \nu) \leq f_0(\tilde{x})$ and

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu) \leq L(\tilde{x}, \lambda, \nu) \leq f_0(\tilde{x})$$

which holds for every feasible $\tilde{x}$.
Maximizing $g(\lambda, \nu)$ to find the best lower bound: the **dual problem**.

Primal and Dual Problems

- Original problem is equivalent to

$$\min_{x} \quad (\sup_{\lambda \succeq 0, \nu} L(x, \lambda, \nu))$$

- Dual problem is switching the min and max

$$\max_{\lambda \succeq 0, \nu} \quad (\inf_{x} L(x, \lambda, \nu))$$

  - The dual function $g(\lambda, \nu) = \inf_x L(x, \lambda, \nu)$ is concave
  - The dual problem is a convex optimization (maximization of a concave function and linear constraints).
  - The optimal value is denoted $d^\star \leq p^\star$
  - Find the best lower bound on $p^\star$ from the dual function.

Interpretation from the Min-max inequality

For any function L of two variables

$$\min_x \max_y L(x,y) \geq \max_y \min_x L(x,y)$$

Proof

$$\max_y L(x,y) \geq L(x,y) \geq \min_x L(x,y) \quad \forall y, x$$

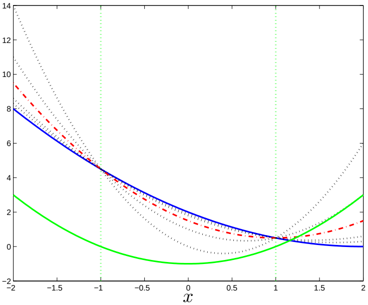
Then taking the minimum over x on the RHS, then the maximum of the LHS.

Geometric Look

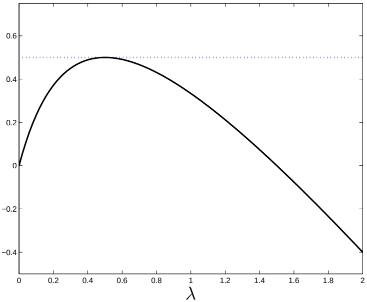
$$\min_{x} \quad \frac{1}{2}(x-2)^2, \quad \text{s.t.} \quad x^2 \le 1$$

Langrangian $L(x,\lambda) = \dfrac{1}{2}(x-2)^2 + \lambda(x^2-1)$

Dual function $g(\lambda) = \inf_{x} \dfrac{1}{2}(x-2)^2 + \lambda(x^2-1)$



True function (blue), constraint (green), $\mathcal{L}(x,\lambda)$ for different λ (dotted)



Dual function $g(\lambda)$ (black), primal optimal (dotted blue)

Geometric Look 2

$$\min_x \quad \frac{1}{2}(x-2)^2, \quad \text{s.t.} \quad x^2 \leq 1$$

Langrangian $L(x,\lambda) = \dfrac{1}{2}(x-2)^2 + \lambda(x^2-1)$

Dual function $g(\lambda) = \inf\limits_x \dfrac{1}{2}(x-2)^2 + \lambda(x^2-1)$

Set $\frac{d}{dx}L(x,\lambda) = x - 2 + 2\lambda x = 0$ to find $x = \frac{2}{2\lambda+1}$

Put x into $g(\lambda) = \frac{\lambda(3-2\lambda)}{2\lambda+1}$

Uses of the Dual

- **Weak duality** always holds (even for nonconvex problems):

$$d^\star \leq p^\star$$

- **Certificate of optimality** (a.k.a. duality gap)
  If we have feasible x and know the dual $g(\lambda, \nu)$, then

$$g(\lambda, \nu) \leq f_0(x^\star) \leq f_0(x)$$

$$f_0(x) - f_0(x^\star) \leq f_0(x) - g(\lambda, \nu).$$

The Greatest Property of the Dual

Theorem For reasonable convex problems,

$$\sup_{\lambda \succeq 0, \nu} g(\lambda, \nu) = f_0(x^\star)$$

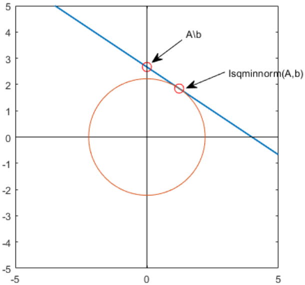
Strong duality means that the duality gap is zero.

- is very desirable (solve a dual problem sometimes easier than the original problem)
- does not hold in general
- usually holds for convex problems
- conditions that guarantee strong duality in convex problems are called constraint qualifications

Example: Least-Norm Solution of Linear Equations I

- Solve an underdetermined linear equations
  $Ax = b$ with $b \in \mathbb{R}^m$, $A \in \mathbb{R}^{m \times n}$, $m < n$.
- A is a wide matrix (more columns than rows). The system may have many solutions.
- Find a solution with least norm, i.e.,

$$\min_x \quad x^T x \quad \text{s.t.} \quad Ax = b$$



Example $2x_1 + 3x_2 = 8$

Example: Least-Norm Solution of Linear Equations II

- **Lagrangian**

$$L(x, \nu) = x^T x + \nu^T (Ax - b)$$

- To find the **dual function**, we solve an unconstrained minimization of the Lagrangian.
  Set the gradient equal to zero

$$\nabla_x L(x, \nu) = 2x + A^T \nu = 0 \quad \implies x^\star = -(1/2)A^T \nu$$

  and plug the solution into $L(x, \nu)$

$$g(\nu) = L(-(1/2)A^T \nu, \nu) = \frac{-1}{4} \nu^T A A^T \nu - b^T \nu$$

- **Lower bound**

$$p^\star \geq -\frac{1}{4} \nu^T A A^T \nu - b^T \nu, \quad \forall \nu$$

Example: Least-Norm Solution of Linear Equations III

- **Dual problem** is the Quadratic Programming

$$\max_{\nu} \quad -\frac{1}{4}\nu^T AA^T \nu - b^T \nu$$

which has a solution $\nu^\star = -2(AA^T)^{-1}b$ and

$$d^\star = b^T(AA^T)^{-1}b$$

- The least-norm solution $x^\star = A^T(AA^T)^{-1}b$ and

$$p^\star = b^T(AA^T)^{-1}b = d^\star$$

- **Question**: solution with weighted energy, i.e.

$$\min_{x} \quad \sum_n w_n x_n^2 \quad \text{s.t.} \quad Ax = b$$

where $w_n > 0$.

Example: Least-Norm Solution of Linear Equations IV

Hint: reparameterize $z_n = \sqrt{w_n} x_n$

$$\min_z \quad z^T z \qquad \text{s.t.} \quad (A\,\mathrm{diag}(1/\sqrt{w}))z = b$$

$$x^\star = \mathrm{diag}(1/w) A^T (A\,\mathrm{diag}(1/w) A^T)^{-1} b$$

Example: Sparse Solution of Linear Equations I

- Find a sparsest solution, i.e.,

$$\min_{x} \quad \|x\|_1 \quad \text{s.t.} \quad Ax = b$$

Solve an alternating problem

$$\min_{x, z > 0} \quad \sum_{n} z_n + \frac{x_n^2}{z_n}$$
$$\text{s.t.} \quad Ax = b$$

We can sequentially update z and x.

- The optimal $z^\star = |x|$

$$z_n + \frac{x_n^2}{z_n} \geq 2|x_n|$$

Example: Sparse Solution of Linear Equations II

- x is solution to a QP with linear constraint

$$\begin{aligned} \min_x \quad & \sum_n \frac{x_n^2}{z_n} \\ \text{s.t.} \quad & Ax = b \end{aligned}$$

$$x^\star = \operatorname{diag}(z) A^T (A \operatorname{diag}(z) A^T)^{-1} b$$

- In summary, we have the following update rule

$$x^\star \leftarrow \operatorname{diag}(|x|) A^T (A \operatorname{diag}(|x|) A^T)^{-1} b$$

Example: Linearly constrained least squares

$$\min_x \quad \frac{1}{2}\|Ax - b\|^2 \quad \text{s.t.} \quad Bx = d$$

- **Lagrangian**

$$L(x, \nu) = \frac{1}{2}\|Ax - b\|^2 + \nu^T(Bx - d)$$

- **Dual function**

Take infimum by setting gradient to zero

$$\nabla_x L(x, \nu) = A^T Ax - A^T b + B^T \nu = 0$$

to give $x^\star = (A^T A)^{-1}(A^T b - B^T \nu)$

Dual function is a simple unconstrained quadratic problem

$$\begin{aligned} g(\nu) &= \inf_x L(x, \nu) \\ &= \frac{1}{2}\|A(A^T A)^{-1}(A^T b - B^T \nu) - b\|^2 + \nu^T B(A^T A)^{-1}(A^T b - B^T \nu) - d^T \nu \end{aligned}$$

Example: Quadratically constrained least squares I

$$\min_x \frac{1}{2}\|Ax - b\|_2^2 \quad \text{s.t.} \quad \frac{1}{2}\|x\|_2^2 \le c$$

where $A \in \mathbb{R}^{n\times m}$.

- **Lagrangian** ($\lambda \ge 0$):

$$L(x,\lambda) = \frac{1}{2}\|Ax - b\|_2^2 + \frac{1}{2}\lambda(\|x\|_2^2 - 2c)$$

- **Dual function** Take infimum:

$$\nabla_x L(x,\lambda) = A^T Ax - A^T b + \lambda x = 0$$

at $x^\star = (A^T A + \lambda I)^{-1} A^T b$ and the dual function

$$\begin{aligned} g(\lambda) &= \inf_x L(x,\lambda) \\ &= \frac{1}{2}\|A(A^T A + \lambda I)^{-1}A^T b - b\|_2^2 + \frac{\lambda}{2}\|((A^T A + \lambda I)^{-1} A^T b)\|_2^2 - \lambda^T c \end{aligned}$$

One variable dual problem!
Can we solve the dual problem?

Example: Quadratically constrained least squares II

- If $\lambda = 0$, then $x^\star = x_{LS} = (A^T A)^{-1} A^T b$. Optimal if $\|x_{LS}\|_2^2 \le 2c$.
- Otherwise when $\|x_{LS}\|_2^2 > 2c$, $\lambda > 0$.
  Denote SVD of $A = U\Sigma V^T$, where $\Sigma = \text{diag}(\sigma)$, and $z = U^T b$.

$$g(\lambda) = \frac{\lambda}{2} \sum_i \frac{z_i^2}{\sigma_i^2 + \lambda} - \lambda c$$

$$g'(\lambda) = \frac{1}{2} \sum_i \frac{z_i^2 \sigma_i^2}{(\sigma_i^2 + \lambda)^2} - c$$

  Note that
  $g'(\lambda) = 0$ has a unique solution $\lambda^\star > 0$.
  $g'(0) = \frac{1}{2} b^T (AA^T)^{-1} b - c = \frac{1}{2} \|x_{LS}\|^2 - c > 0$.
  $g'(\infty) \to -c$

Example: Linear Programming I

- Consider the problem

$$\begin{array}{ll} \underset{x}{\text{minimize}} & c^T x \\ \text{subject to} & Ax = b, \quad x \geq 0. \end{array}$$

- The Lagrangian is

$$\begin{aligned} L(x, \lambda, \nu) &= c^T x + \nu^T (Ax - b) - \lambda^T x \\ &= \left(c + A^T \nu - \lambda\right)^T x - b^T \nu. \end{aligned}$$

- L is a linear function of x and it is unbounded if the term multiplying x is nonzero.

Example: Linear Programming II

- Hence, the dual function is

$$g(\lambda, \nu) = \inf_x L(x, \lambda, \nu) = \begin{cases} -b^T \nu & c + A^T \nu - \lambda = 0 \\ -\infty & \text{otherwise.} \end{cases}$$

- The function g is a concave function of (λ, ν) as it is linear on an affine domain.

- From the lower bound property, we have

$$p^\star \geq -b^T \nu \quad \text{if } c + A^T \nu \geq 0.$$

- The dual problem is the LP

$$\begin{array}{ll} \underset{\nu}{\text{maximize}} & -b^T \nu \\ \text{subject to} & c + A^T \nu \geq 0. \end{array}$$

Example: Two-Way Partitioning I

Partitioning problem

- Given a set of n elements $1, ..., n$.
- Denote by $W_{i,j} \in \mathbb{S}^n$ the cost of having the elements i and j in the same set. $W_{i,i} = 0$ for $i = 1, \ldots, n$.
- We need to partition into two sets while minimizing the total cost.
- Let $x = [x_1, x_2, \ldots, x_n]^T \in \{\pm 1\}^n$, which corresponds to the partition

$$\{1, \ldots, n\} = \{i \,|\, x_i = 1\} \cup \{i \,|\, x_i = -1\}.$$

Example: Two-Way Partitioning II

- Consider the problem

$$\begin{array}{ll} \underset{x}{\text{minimize}} & x^T W x \\ \text{subject to} & x_i^2 = 1, \quad i = 1, \ldots, n. \end{array}$$

- It is a nonconvex problem (quadratic equality constraints). The feasible set contains 2^n discrete points.
- The Lagrangian is

$$\begin{aligned} L(x, \nu) &= x^T W x + \sum_{i=1}^{n} \nu_i \left(x_i^2 - 1 \right) \\ &= x^T \left(W + \text{diag}(\nu) \right) x - 1^T \nu. \end{aligned}$$

- L is a quadratic function of x and it is unbounded if the matrix $W + \text{diag}(\nu)$ has a negative eigenvalue.

Example: Two-Way Partitioning III

- Hence, the dual function is

$$g(\nu) = \inf_x L(x,\nu) = \begin{cases} -1^T\nu & W + \operatorname{diag}(\nu) \succeq 0 \\ -\infty & \text{otherwise.} \end{cases}$$

- From the lower bound property, we have

$$p^\star \geq -1^T\nu \quad \text{if } W + \operatorname{diag}(\nu) \succeq 0.$$

- As an example, if we choose $\nu = -\lambda_{\min}(W)\,1$, we get the bound

$$p^\star \geq n\lambda_{\min}(W).$$

- The dual problem is the SDP

$$\begin{array}{ll} \underset{\nu}{\text{maximize}} & -1^T\nu \\ \text{subject to} & W + \operatorname{diag}(\nu) \succeq 0. \end{array}$$

Example: Two-Way Partitioning IV

Bi-dual of BQP.
Lagrangian

$$\begin{aligned} L(\nu, Z) &= 1^T\nu - \mathsf{trace}(Z(W + \mathsf{diag}(\nu))) \\ &= \nu^T 1 - \mathsf{trace}(ZW) - \nu^T \mathsf{diag}(Z) \\ &= \nu^T(1 - \mathsf{diag}(Z)) - \mathsf{trace}(ZW) \end{aligned}$$

The dual function is

$$g(Z) = \inf_{\nu} L(\nu, Z) = \begin{cases} -\mathsf{trace}(ZW) & \mathsf{diag}(Z) = 1 \\ -\infty \quad , & otherwise \end{cases}$$

The dual problem can be expressed

$$\begin{aligned} \min \quad & \mathsf{trace}(ZW) \\ \text{s.t.} \quad & Z \succeq 0 \\ & Z_{ii} = 1 \end{aligned}$$

Example: Two-Way Partitioning V

$Z \succeq 0$ is a convex relaxation of xx^T in the original primal problem.

```
cvx_begin sdp
    variable nu(N)
    maximize ( -sum(nu) )
    subject to
        W + diag(nu) ≥ 0;
cvx_end
```

```
cvx_begin sdp
    variable Z(N,N) symmetric
    minimize ( trace(W*Z) )
    subject to
     diag(Z) == 1;
       Z ≥ 0;
cvx_end
```

Lagrange dual and conjugate function

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & Ax \preceq b, \quad Cx = d \end{array}$$

dual function

$$\begin{aligned} g(\lambda, \nu) &= \inf_{x \in \mathbf{dom}\, f_0} \left(f_0(x) + (A^T\lambda + C^T\nu)^T x - b^T\lambda - d^T\nu \right) \\ &= -f_0^*(-A^T\lambda - C^T\nu) - b^T\lambda - d^T\nu \end{aligned}$$

- recall definition of conjugate $f^*(y) = \sup_{x \in \mathbf{dom}\, f}(y^T x - f(x))$
- simplifies derivation of dual if conjugate of f_0 is known

example: entropy maximization

$$f_0(x) = \sum_{i=1}^n x_i \log x_i, \qquad f_0^*(y) = \sum_{i=1}^n e^{y_i - 1}$$

Example: Entropy maximization I

$$\begin{aligned} \min \quad & \sum_i x_i \log x_i \\ \text{s.t.} \quad & \sum_i x_i = 1, \quad x \in \mathbb{R}^n_+. \end{aligned}$$

- $f(x) = x \log x$ and its conjugate function $f^\star(y) = e^{y-1}$
- **dual function**

$$\begin{aligned} g(\lambda, \nu) &= \inf_x (\sum_i x_i \log(x_i) - \lambda^T x + \nu(1_n^T x - 1) \\ &= -\nu - f^\star(\lambda - 1_n \nu) \\ &= -\nu - \sum_{i=1}^n e^{\lambda_i - \nu - 1} \end{aligned}$$

- Can we solve the dual problem $\max g(\lambda, \nu)$?

Example: Entropy maximization II

- The dual function $g(\lambda, \nu)$ attains maximum when $\lambda = 0$ and $\nu = 1 - \log(n)$
- Dual optimal $d^\star = g(\lambda^\star, \nu^\star) = -\log(n)$
- Primal optimal $p^\star = \log(n)$ with $x_i^\star = \frac{1}{n}$

```
cvx_begin
variables x(n,1)
dual variable lambda       % dual variable for x>0
dual variable nu           % dual variable for sum(x) = 1
minimize -sum(entr(x));    % entr(x) = -x log(x)
subject to
   nu: sum(x)==1;
   lambda: x>0;
cvx_end
```

Example: Convex piecewise-linear minimization I

$$\min \max_{i=1,\dots,m} (a_i^T x + b_i) \qquad x \in \mathbb{R}^n$$

- Equivalent problem

$$\begin{aligned} \min \max_{i=1,\dots,m} \quad & y_i \\ \text{s.t.} \quad & a_i x + b_i = y_i, \quad i = 1, \dots, m, \end{aligned}$$

  with variables $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$

- **Dual function**

$$g(\lambda) = \inf_{x,y} \left(\max_{i=1,\dots,m} y_i + \sum_{i=1}^{m} \lambda_i (a_i^T x + b_i - y_i) \right)$$

  The infimum over x is finite only if $\sum_{i=1}^{m} \lambda_i a_i = 0$.

Example: Convex piecewise-linear minimization II

- Minimize over y

$$\inf_y \left(\max_i y_i - \lambda^T y \right) = \begin{cases} 0, & \lambda \geq 0, \mathbf{1}^T \lambda = 1, \\ -\infty, & \text{otherwise} \end{cases}$$

Note: For norm $\|y\|_\infty$, its dual norm $\|y\|^\star = \|y\|_1$, its conjugate is indicator of unit ball for dual norm.

$$f^\star(y) = \sup_x (y^T x - \|x\|_\infty) = \begin{cases} 0, & \|y\|^\star \leq 1 \\ +\infty & \|y\|^\star > 1 \end{cases}$$

- If $\lambda \geq 0, \mathbf{1}^T \lambda = 1$, then $\max_i y_i - \lambda^T y = \sum \lambda_i (\max_i y_i - y_i) \geq 0$
- If $\mathbf{1}^T \lambda \neq 1$, select $y = t\mathbf{1}$
- If exists $\lambda_i < 0$, select $y = te_i$,

Example: Convex piecewise-linear minimization III

- Summing up, we have the dual function

$$g(\lambda) = \begin{cases} b^T\lambda & A^T\lambda = 0, \mathbf{1}^T\lambda = 1, \lambda \succeq 0, \\ -\infty & \text{otherwise.} \end{cases}$$

The dual problem

$$\begin{aligned} \max \quad & b^T\lambda \\ \text{s.t.} \quad & A^T\lambda = 0 \\ & \mathbf{1}^T\lambda = 1 \\ & \lambda \succeq 0. \end{aligned}$$

Example: Convex piecewise-linear minimization IV

- Another representation

$$\begin{aligned} \min \quad & t \\ \text{s.t.} \quad & Ax + b \le t\mathbf{1}. \end{aligned}$$

Obtain an identical dual problem

$$\begin{aligned} \max \quad & b^T\lambda \\ \text{s.t.} \quad & A^T\lambda = 0, \mathbf{1}^T\lambda = 1, \lambda^T \succeq 0 \end{aligned}$$

Slater's Constraint qualification I

$$\begin{array}{rl} \min & f(x) \\ \text{subject to} & g_j(x) \leq 0, \quad j = 1, \ldots, m \\ & x \in X \end{array}$$

- If there exists a strictly feasible $\bar{x} \in X \bigcap \text{dom} f$ such that

$$g_j(\bar{x}) < 0 \quad \text{for all } j = 1, \ldots, m.$$

  It guarantees the dual optimum is attained and there is no gap, i.e., $p^\star = d^\star$.

Slater's Constraint qualification II

Proof

- Define a set $V = \{(u, w) \,|\, g(x) \leq u, f(x) \leq w, x \in X\} \subset \mathbb{R}^m \times R$, where $g(x) = [g_1(x), \ldots, g_m(x)]$.
  - The set V is convex by the convexity of f and g_j and X.
- First, the point $(\mathbf{0}, p^\star)$ is not in the interior of V
  - Otherwise, if $(\mathbf{0}, p^\star) \in \text{int}V$, there exists $\varepsilon > 0$ such that $(\mathbf{0}, p^\star - \varepsilon) \in V$, i.e., contradicting the optimality of $p^\star$.
- The point $(\mathbf{0}, p^\star) \in \text{bd}V$ or not in V.
  - There exists a (supporting) hyperplane which separates V and the point $(\mathbf{0}, p^\star)$, i.e., $\exists(\lambda, \nu) \in \mathbb{R}^m \times \mathbb{R}$ with $(\lambda, \nu) \neq 0$ such that

$$\lambda^T u + \nu w \geq \lambda^T \mathbf{0} + \nu p^\star = \nu p^\star \quad \text{for all } (u, w) \in V \tag{1}$$

    Imply that $\lambda \geq 0$ and $\nu \geq 0$. Otherwise, $\lambda^T u + \nu w$ is unbounded.

Slater's Constraint qualification III

- We will prove $\nu > 0$ by contradiction.
  Suppose $\nu = 0$, then $\lambda \neq 0$ and the condition in (**??**) reduces to

$$\inf_{(u,w)\in V} \quad \lambda^T u = 0$$

  In addition, since $\lambda \geq 0$ and $\lambda \neq 0$

$$\inf_{(u,w)\in V} \quad \lambda^T u = \inf_{x\in X} \quad \lambda^T g(x) \leq \lambda^T g(\bar{x}) < 0$$

  Contradiction!
- Dividing both sides of (**??**) with ν, we obtain

$$\inf_{(u,w)\in V} \quad \tilde{\lambda}^T u + w \geq p^\star$$

  where $\tilde{\lambda} = \frac{\lambda}{\nu} \geq 0$. It means

$$\inf_{x\in X} \quad f(x) + \tilde{\lambda}^T g(x) \geq p^\star$$

  or $d^\star \geq p^\star$. By the weak duality $d^\star \leq p^\star$, it follows that $d^\star = p^\star$ and $\tilde{\lambda}$ is the dual optimal solution.

Example

$$\min_{x,y} \quad \exp(-x), \quad \text{s.t.} \quad x^2/y \leq 0$$

over the domain $\mathcal{D} = \{(x,y) \mid y > 0\}$

- Only feasible $x = 0$. So $p^\star = 1$
- **Langragian** $L(x,y,\lambda) = \exp(-x) + \lambda x^2/y$
  So dual function is

$$g(\lambda) = \inf_{x,y} \quad \exp(-x) + \lambda x^2/y = \begin{cases} 0 & \lambda \geq 0 \\ -\infty & \lambda < 0 \end{cases}$$

- **Dual problem**

$$d^\star = \max_{\lambda} 0, \quad \text{s.t.} \quad \lambda \geq 0$$

  Thus $d^\star = 0$ and the dual gap $p^\star - d^\star = 1$.
  Here we has no strictly feasible solution.

Trust region subproblem I

$$\begin{aligned} \min_x \quad & x^T A x + 2b^T x \\ \text{s.t.} \quad & x^T x \le 1 \end{aligned}$$

where $A \not\succeq 0$, i.e., the problem is non-convex.

- **dual function** $g(\lambda) = \inf_x \quad (x^T(A + \lambda I)x + 2b^T x - \lambda)$
  - unbounded below if $A + \lambda I \not\succeq 0$ or if $A + \lambda I \succeq 0$ and $b \notin \mathcal{R}(A + \lambda I)$ Why?
  - minimized by $x^\star = -(A + \lambda I)^\dagger b$, and $g(\lambda) = -b^T(A + \lambda I)^\dagger b - \lambda$
- **Lagrange dual problem** (convex)

$$\begin{aligned} \max \quad & -b^T(A + \lambda I)^\dagger b - \lambda \\ \text{s.t.} \quad & A + \lambda I \succeq 0, \quad b \in \mathcal{R}(A + \lambda I) \end{aligned}$$

where $(A + \lambda I)^\dagger$ is the pseudo-inverse of $A + \lambda I$.

Trust region subproblem II

- Another form

$$\begin{aligned} \max \quad & -\sum_{i=1}^{n} \frac{(q_i^T b)^2}{\lambda_1 + \lambda} - \lambda \\ \text{s.t.} \quad & \lambda \geq -\lambda_{min}(A) \end{aligned}$$

where q_i and λ_i are eigenvectors and eigenvalues of A.
Strong duality although primal problem is not convex.

```
[Q,sigma] = eig(A);
sigma = diag(sigma);
c = (Q'*b).^2;

cvx_begin
variable lambda;
minimize (c'*inv_pos(lambda + sigma)+lambda)
subject to
    lambda ≥ - min(sigma);
cvx_end
```

Support vector machine I

$$\begin{aligned} \min_{x,\xi} \quad & \frac{1}{2}\|x\|_2^2 + C\sum_i \xi_i \\ \text{s.t.} \quad & Ax \le \mathbf{1} - \xi, \quad \xi \ge 0 \end{aligned}$$

Lagrangian function

$$L(x,\xi,\lambda,\nu) = \frac{1}{2}\|x\|_2^2 + C\,\mathbf{1}^T\xi - \lambda^T(Ax - \mathbf{1} + \xi) - \nu^T\xi$$

Dual function

$$\begin{aligned} g(\lambda,\nu) &= \inf L(x,\xi,\lambda,\nu) \\ &= \begin{cases} \lambda^T\mathbf{1} - \frac{1}{2}\|A^T\lambda\|_2^2, & \lambda+\nu = C\mathbf{1} \\ +\infty & otherwise \end{cases} \\ d^\star &= \max_{\lambda\ge 0,\nu\ge 0} \quad g(\lambda,\nu) \end{aligned}$$

Support vector machine II

Question Can we eliminate ν from above problem?

$$d^\star = \max_{\lambda \geq 0} g(\lambda, \nu) = \max_{\lambda} \lambda^T \mathbf{1} - \frac{1}{2} \lambda^T A^T A \lambda, \quad 0 \leq \lambda \leq C\mathbf{1}$$

LASSO I

$$\min_x \quad \frac{1}{2}\|y - Ax\|_2^2 + \gamma\|x\|_1$$

Reformulate

$$\begin{aligned} \min_x \quad & \frac{1}{2}\|y - z\|_2^2 + \gamma\|x\|_1 \\ \text{s.t.} \quad & z = Ax \end{aligned}$$

Lagrangian function

$$L(z, x, \nu) = f(z) + r(x) + \nu^T(z - Ax)$$

where

$$f(z) = \frac{1}{2}\|y - z\|_2^2 \qquad r(x) = \gamma\|x\|_1$$

Dual function

LASSO II

$$
\begin{aligned}
g(\nu) &= \inf_x f(z) + r(x) + \nu^T(z - Ax) \\
&= -\sup_z(-z^T\nu - f(z)) - \sup_x(x^T A^T \nu - r(x)) \\
&= -f^\star(-\nu) - r^\star(A^T\nu)
\end{aligned}
$$

where

$$
f(z) = \frac{1}{2}\|y - z\|_2^2 \qquad\qquad f^\star(\nu) = \frac{1}{2}\nu^T\nu + y^T\nu
$$

$$
r(x) = \gamma\|x\|_1 \qquad\qquad r^\star(\nu) = \begin{cases} 0 & \|\nu\|_\infty \leq \gamma \\ \infty & \text{otherwise} \end{cases}
$$

Dual problem

$$
\max \quad \frac{-1}{2}\nu^T\nu + y^T\nu \quad \text{s.t.} \quad \|A^T\nu\|_\infty \leq \gamma\,.
$$

LASSO III

How to find optimal $x^\star$
Denote $b = A^T\nu^\star$. Since $\nu^\star$ is solution to the dual problem, $\|b\|_\infty \leq \gamma$.
z is solution of the LS problem

$$\inf_z \frac{1}{2}\|y - z\|_2^2 + \nu^T z$$

has the optimal solution $z^\star = y - \nu$ and the minimization problem

$$\inf_x \ \gamma\|x\|_1 - b^T x = \sum_n |x_n|(\gamma - b_n \operatorname{sign}(x_n))$$

has optimal solution

- $x_n^\star = 0$ if $|b_n| < \gamma$
- and $x_{\mathcal{I}}$ is solution to a linear system $A(:, \mathcal{I})x_{\mathcal{I}} = z^\star$ where , $\mathcal{I} = \{n : |b_n| = \gamma\}$

LASSO IV

In practice, there might be no $|b_n| = \gamma$ but close to γ. Denote $\mathcal{I} = \{n : |b_n - \gamma| \le \varepsilon\}$ where $\varepsilon = 10^{-4}$, then $x^\star_{n \notin I} = 0$ and $x^\star_I = A^\dagger_I z^\star$.

```
% Solve the dual problem
cvx_begin
variable nu(m)
minimize(0.5*sum_square(nu)-y'*nu)
subject to  norm(A'*nu,Inf) ≤ gamma
cvx_end

%Find the Optimal  z
z = y-nu;

% Find optimal x
x = zeros(n,1);
b = A'*nu;
nnzix = find((gamma-abs(b))≤1e-4);
x(nnzix) = A(:,nnzix)\z;
```

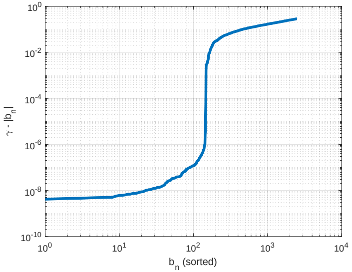


Figure: In this example, there are 147 $b_n \approx \gamma$, i.e., 147 nonzeros x_n.

LASSO V

Alternative method: Check the optimal values $p^\star$ with different selections of nonzeros x_n, and compare those with the dual optimal.
Choose the solution with the smallest gap $p^\star - d^\star$.

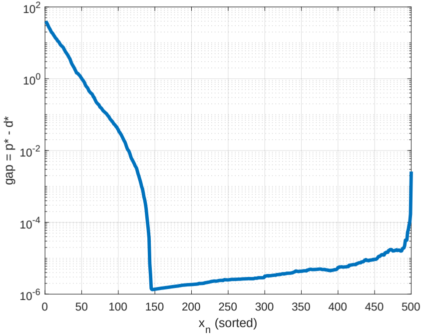


Figure: Gap $p^\star - d^\star$ is smallest with 147 nonzeros x_n.

(Smooth) Fused Lasso I

Minimize the least squares error measure with the additional requirements: sparse and "smooth" solution

$$\min_{x} \quad \frac{1}{2}\|y - Ax\|_2^2 + \gamma_1\|x\|_1 + \gamma_2\|Dx\|_1$$

where D can be derivative operator or inverse of wavelet transform

(Smooth) Fused Lasso II

$$\min_x \quad \frac{1}{2}\|y - Ax\|_2^2 + \gamma_1 \|Fx\|_1$$

where $F = [I; \frac{\gamma_2}{\gamma_1} D]$.

Dual Problem for Regularized optimization I

Consider a general Regularized optimization problem

$$\inf_{x \in X} \quad f(x) + r(Ax) \qquad \text{s.t.} \quad Ax \in Y$$

Introduce new variable $z = Ax$

$$\inf_{x \in X, z \in Y} \quad f(x) + r(z) \qquad \text{s.t.} \quad z = Ax\,.$$

Lagragian function

$$L(x, z, \nu) = f(x) + r(z) + \nu^T(Ax - z), \quad x \in X, z \in Y\,.$$

Dual Problem for Regularized optimization II

Associated dual function

$$\begin{aligned} g(\nu) &= \inf_{x\in X, z\in Y} \quad f(x) + \nu^T Ax + \inf_{z\in Y} \quad r(z) - \nu^T z \\ &= -\sup_{x\in X}\{-x^T A^T \nu - f(x)\} - \sup_{z\in Y}\{z^T \nu - r(z)\} \\ &= -f^\star(-A^T\nu) - r^\star(\nu) \end{aligned}$$

Feasible descent direction I

Consider the problem

$$\min \quad f(x) \quad \text{s.t.} \quad x \in C,$$

where f is continuously differentiable over $C \subseteq \mathbb{R}^n$.

- A vector d is a **Feasible descent direction** at $x \in C$ if $\nabla f(x)^T d < 0$ and exist $\varepsilon > 0$ s.t $x + td \in C$ for all $t \in [0, \varepsilon]$.
- If $x^\star$ is a local optimal solution, then there are no feasible descent directions.
  Hint Consider the optimality condition $\nabla f(x^\star)^T (y - x^\star) \geq 0$ for all $y \in C, \ldots$.
  Or consider the function $g(t) = f(x^\star + td)$, $\ldots$,

Necessary Condition I

Let $x^\star$ be a local minimum of the problem

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & f(x) \\ \text{s.t} \quad & g_i(x) \le 0, \quad i = 1, 2, \ldots, m \end{aligned}$$

where f, $g_1, \ldots, g_m$ are continuously differentiable functions.

- There exist multipliers $\lambda_0, \lambda_1, \ldots, \lambda_m \ge 0$, not all zeros, such that

$$\lambda_0 \nabla f(x^\star) + \sum_{i=1}^{m} \lambda_i \nabla g_i(x^\star) = 0, \quad \lambda_i g_i(x^\star) = 0, \quad i = 1, 2, \ldots, m.$$

Necessary Condition II

Proof

Define a set of active constraints at $x^\star$, $I(x^\star) = \{i : g_i(x^\star) = 0\}$.
Since $x^\star$ is a local minima, there does not exist a vector d such that

$$\nabla f(x^\star)^T d < 0\,, \qquad \nabla g_i(x^\star)^T d < 0, \quad i \in I(x^\star)\,.$$

(Why???)

Necessary Condition III

Brief proof:

- Suppose there is such a vector d, we will prove that d is a feasible descent direction at $x^\star$
- Then there exists $\varepsilon > 0$ such that

$$f(x^\star+td) < f(x^\star), \quad g_i(x^\star+td) < g_i(x^\star) = 0, \text{for any}\, t \in (0,\varepsilon), i \in I(x^\star)$$

- For any $i \notin I(x^\star)$, $g_i(x^\star) < 0$, by the continuity of g_i, there exists $\varepsilon_2 > 0$ such that $g_i(x^\star + td) < 0$ for any $t \in (0, \varepsilon_2)$.
- Hence for all $t \in (0, \min(\varepsilon, \varepsilon_2))$

$$f(x^\star + td) < f(x^\star), \qquad g_i(x^\star + td) < g_i(x^\star) = 0, \qquad i = 1, \ldots, m,$$

i.e., d is a feasible descent direction at $x^\star$.
Contradiction to the local optimality of $x^\star$.

Necessary Condition IV

(continue the main proof)
i.e., the linear system $Ad < 0$ is infeasible
where

$$A = \begin{bmatrix} \nabla f(x^\star)^T \\ \nabla g_{i_1}(x^\star)^T \\ \vdots \\ \nabla g_{i_k}(x^\star)^T \end{bmatrix}, \qquad i_k \in I(x^\star)$$

Following the Farkas lemma, the dual system has a nonnegative solution

$$A^T\lambda = 0, \quad \lambda = (\lambda_0, \lambda_{i_1}, \ldots, \lambda_{i_k}) \geq 0\,.$$

Next, for $i \notin I(x^\star)$, define $\lambda_i = 0$. We finally obtain the proof.

Complementary slackness

assume strong duality holds, $x^\star$ is primal optimal, $(\lambda^\star, \nu^\star)$ is dual optimal

$$
\begin{array}{rcl}
f_0(x^\star) = g(\lambda^\star, \nu^\star) & = & \inf_x \left(f_0(x) + \sum_{i=1}^{m} \lambda_i^\star f_i(x) + \sum_{i=1}^{p} \nu_i^\star h_i(x) \right) \\
& \leq & f_0(x^\star) + \sum_{i=1}^{m} \lambda_i^\star f_i(x^\star) + \sum_{i=1}^{p} \nu_i^\star h_i(x^\star) \\
& \leq & f_0(x^\star)
\end{array}
$$

hence, the two inequalities hold with equality

- $x^\star$ minimizes $L(x, \lambda^\star, \nu^\star)$
- $\lambda_i^\star f_i(x^\star) = 0$ for $i = 1, \ldots, m$ (known as complementary slackness):

$$
\lambda_i^\star > 0 \Longrightarrow f_i(x^\star) = 0, \qquad f_i(x^\star) < 0 \Longrightarrow \lambda_i^\star = 0
$$

Karush-Kuhn-Tucker Optimality conditions I

KKT conditions (for differentiable f_i, h_i):

1. primal feasibility: $f_i(x) \leq 0,\ i = 1, \dots, m,\ h_i(x) = 0,\ i = 1, \dots, p$
2. dual feasibility: $\lambda \geq 0$
3. complementary slackness: $\lambda_i^\star f_i(x^\star) = 0 \text{ for } i = 1, \dots, m$
4. zero gradient of Lagrangian with respect to x:

$$\nabla f_0(x) + \sum_{i=1}^{m} \lambda_i \nabla f_i(x) + \sum_{i=1}^{p} \nu_i \nabla h_i(x) = 0$$

Karush-Kuhn-Tucker Optimality conditions II

- We already known that if strong duality holds and x, λ, ν are optimal, then they must satisfy the KKT conditions.
- What about the opposite statement?
- If x, λ, ν satisfy the KKT conditions for a convex problem, then they are optimal.

Proof. From complementary slackness, $f_0(x) = L(x, \lambda, \nu)$ and, from 4th KKT condition and convexity, $g(\lambda, \nu) = L(x, \lambda, \nu)$. Hence, $f_0(x) = g(\lambda, \nu)$. $\square$

Theorem. *If a problem is convex and Slater's condition is satisfied, then x is optimal if and only if there exists λ, ν that satisfy the KKT conditions.*

Examples of Using KKT Conditions

$$\begin{array}{ll} \text{minimize} & (1/2)x^T P x + q^T x + r \\ \text{subject to} & Ax = b \end{array}$$

with $P \in \mathbb{S}^n_{++}$

KKT conditions:

$$Ax^\star = b, \quad Px^\star + q + A^T \nu^\star = 0$$

KKT conditions rewritten:

$$\left[\begin{array}{cc} P & A^T \\ A & 0 \end{array}\right] \left[\begin{array}{c} x^\star \\ \nu^\star \end{array}\right] = \left[\begin{array}{c} -q \\ b \end{array}\right]$$

optimization problem $\Leftrightarrow$ solving linear equations

Examples of Using KKT Conditions

$$\min_x \frac{1}{2}\|Ax - b\|^2 \quad \text{s.t.} \quad \frac{1}{2}\|x\|^2 \leq c.$$

where $A \in \mathbb{R}^{n\times m}$.
Prove that
If $x_{LS} = (A^T A)^{-1} A^T b$ is not feasible, then $\|x^\star\|^2 = 2c$.

Linear regression with bound constraint

$$\min_x \quad \|x\|^2 \quad \text{s.t.} \quad \|Ax - y\| \le \delta$$

where $A \in \mathbb{R}^{n \times m}$ is a regressor matrix and δ a nonnegative regression bound.
Prove that the minimiser to the above problem is the minimiser to the following problem

$$\min_x \quad \|x\|^2 \qquad \text{s.t.} \quad \|y - Ax\| = \delta\,.$$

Dual via Convex conjugates I

$$\min f(x) \quad \text{s.t. } f_i(x) \leq 0, Ax = b.$$

$$\mathcal{L}(x, \lambda, \nu) \quad := \quad f_0(x) + \sum_i \lambda_i f_i(x) + \nu^T(Ax - b)$$

Dual via Convex conjugates II

$$\min f(x) \quad \text{s.t. } f_i(x) \leq 0, Ax = b.$$

$$\begin{aligned}
\mathcal{L}(x, \lambda, \nu) &:= f_0(x) + \sum_i \lambda_i f_i(x) + \nu^T(Ax - b) \\
g(\lambda, \nu) &= \inf_x \mathcal{L}(x, \lambda, \nu) \\
g(\lambda, \nu) &= -\nu^T b + \inf_x x^T A^T \nu + F(x) \\
F(x) &:= f_0(x) + \sum_i \lambda_i f_i(x)
\end{aligned}$$

Dual via Convex conjugates III

$$\min f(x) \quad \text{s.t. } f_i(x) \leq 0, Ax = b.$$

$$\begin{aligned}
\mathcal{L}(x, \lambda, \nu) &= f_0(x) + \sum_i \lambda_i f_i(x) + \nu^T (Ax - b) \\
g(\lambda, \nu) &= \inf_x \mathcal{L}(x, \lambda, \nu) \\
g(\lambda, \nu) &= -\nu^T b + \inf_x x^T A^T \nu + F(x) \\
F(x) &= f_0(x) + \sum_i \lambda_i f_i(x) \\
g(\lambda, \nu) &= -\nu^T b - \sup_x \langle x, -A^T \nu \rangle - F(x) \\
g(\lambda, \nu) &= -\nu^T b - F^*(-A^T \nu).
\end{aligned}$$

Not so useful! F^* hard to compute.

Dual via Convex conjugates

Introduce new variables!

$$\min f(x) \quad \text{s.t.} \quad \begin{aligned} & f_i(x_i) \leq 0, Ax = b \\ & x_i = z, i = 1, \ldots, m. \end{aligned}$$

Dual via Convex conjugates

$$\min f(x) \quad \text{s.t.} \qquad f_i(x_i) \le 0, Ax = b$$
$$x_i = z, i = 1, \ldots, m.$$

$$\mathcal{L}(x, x_i, z, \lambda, \nu, \pi_i)$$
$$= f(x) + \sum_i \lambda_i f_i(x_i) + \nu^T(Ax - b) + \sum_i \pi_i^T(x_i - z)$$
$$g(\lambda, \nu, \pi_i) = \inf_{x, x_i, z} \mathcal{L}(x, x_i, z, \lambda, \nu, \pi_i)$$

Dual via Convex conjugates

$$\min f(x) \quad \text{s.t.} \qquad f_i(x_i) \le 0, Ax = b$$
$$x_i = z, i = 1, \ldots, m.$$

$$\mathcal{L}(x, x_i z, \lambda, \nu, \pi_i)$$
$$= f(x) + \sum\nolimits_i \lambda_i f_i(x_i) + \nu^T(Ax - b) + \sum\nolimits_i \pi_i^T(x_i - z)$$
$$g(\lambda, \nu, \pi_i) = \inf_{x, x_i, z} \mathcal{L}(x, x_i, z, \lambda, \nu, \pi_i)$$
$$= -\nu^T b + \inf_x f(x) + \nu^T Ax + \inf_z \sum\nolimits_i -\pi_i^T z$$
$$+ \sum\nolimits_i \inf_{x_i} \pi_i^T x_i + \lambda_i f_i(x_i)$$

Dual via Convex conjugates

$$\min f(x) \quad \text{s.t.} \quad f_i(x_i) \le 0, Ax = b$$
$$x_i = z, i = 1, \ldots, m.$$

$$\mathcal{L}(x, x_i, z, \lambda, \nu, \pi_i)$$
$$= f(x) + \sum_i \lambda_i f_i(x_i) + \nu^T (Ax - b) + \sum_i \pi_i^T (x_i - z)$$
$$g(\lambda, \nu, \pi_i) = \inf_{x, x_i, z} \mathcal{L}(x, x_i, z, \lambda, \nu, \pi_i)$$
$$= -\nu^T b + \inf_x f(x) + \nu^T Ax + \inf_z \sum_i -\pi_i^T z$$
$$+ \sum_i \inf_{x_i} \pi_i^T x_i + \lambda_i f_i(x_i)$$
$$= \begin{cases} -\nu^T b - f^*(-A^T \nu) - \sum_i (\lambda_i f_i)^*(-\pi_i) & \text{if } \sum_i \pi_i = 0 \\ -\infty & \text{otherwise.} \end{cases}$$

Variable Splitting

$$\min f(x) + h(x)$$

Variable Splitting

$$\min \quad f(x) + h(x)$$

$$\min_{x,z} \quad f(x) + h(z) \qquad \text{s.t.} \quad x = z$$

Lagrangian

$$L(x, z, \nu) = f(x) + h(z) + \nu^T(x - z)$$

Dual

$$g(\nu) = \inf_{x,z} \quad L(x, z, \nu)$$

$$g(\nu) = -f^\star(-\nu) - h^\star(\nu)$$