1. Convex Sets & Geometry

Definitions & Basic Sets

- **Convex Cone:** Set C s.t. $x \in C \implies \theta x \in C$ for $\theta \geq 0$.
- **Hyperplane:** $\{x \mid a^\top x = b\}$. Affine and convex.
- **Halfspace:** $\{x \mid a^\top x \leq b\}$. Convex.
- **Polyhedron:** Intersection of finite number of halfspaces and hyperplanes. $P = \{x \mid Ax \leq b, Cx = d\}$.
- **Euclidean Ball:** $B(x_c, r) = \{x \mid ||x - x_c||_2 \leq r\}$.
- **Ellipsoid:** $\{x \mid (x - x_c)^\top P^{-1}(x - x_c) \leq 1\}$, where $P \in S^n_{++}$. Alternate representation: $\{x_c + Au \mid ||u||_2 \leq 1\}$.
- **Norm Cone:** $\{(x, t) \mid ||x|| \leq t\}$. For $||\cdot||_2$, this is the Second-Order Cone (SOC) or Lorentz Cone.
- **PSD Cone (S^n_+):** The set of symmetric positive semidefinite matrices $\{X \in S^n \mid z^\top Xz \geq 0 \, \forall z\}$. It is a convex cone.

Operations Preserving Convexity

- **Intersection:** even an uncountable number.
- **Affine:** $f(C)$ and $f^{-1}(C)$ are convex.
- **Perspective:** $P(x, t) = x/t$ for $t > 0$. $P(C)$ and $P^{-1}(C)$ convex.
- **Linear-Fractional Function:** $f(x) = (Ax + b)/(c^\top x + d)$ on $\mathbf{dom} f = \{x \mid c^\top x + d > 0\}$. $f(C)$ convex.

Separation Theorems

- **Separating Hyperplane:** If C and D are disjoint convex sets, there exists $a \neq 0, b$ such that $a^\top x \leq b$ for all $x \in C$ and $a^\top x \geq b$ for all $x \in D$.
- **Strict Separation:** Requires C closed, D compact.
- **Supporting Hyperplane:** If C is convex, then at every boundary point x_0, there exists a supporting hyperplane $a \neq 0$ such that $a^\top x \leq a^\top x_0$ for all $x \in C$.

Generalized Inequalities & Cones

A cone K is **proper** if it is convex, closed, solid (nonempty interior), and pointed (contains no lines).

- **Partial Order:** $x \preceq_K y \iff y - x \in K$.
- **Strict Order:** $x \prec_K y \iff y - x \in \mathbf{int} K$.
- **Dual Cone:** $K^* = \{y \mid x^\top y \geq 0 \text{ for all } x \in K\}$.
- **Self-Dual Cones:** The following cones satisfy $K = K^*$: $\mathbb{R}^n_+$, SOCs, and S^n_+.
- **Properties:** $x \preceq_K y \implies \lambda^\top x \leq \lambda^\top y$ for all $\lambda \succeq_{K^*} 0$. $K^{**} = K$ if K is a closed convex cone.

2. Convex Functions

Definitions & Tests

Function $f : \mathbf{dom} f \to \mathbb{R}$ is convex if $\mathbf{dom} f$ is convex and for all $x, y \in \mathbf{dom} f, 0 \leq \theta \leq 1$: $f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y)$
Strictly convex if inequality is strict for $x \neq y, \theta \in (0, 1)$. **Concave** if $-f$ is convex.

Testing Convexity:

- **Restriction to a Line:** f is convex iff $g(t) = f(x + tv)$ is convex in t for all $x \in \mathbf{dom} f, v \in \mathbb{R}^n$.
- **First-Order Cond:** f differentiable. Convex iff $\mathbf{dom} f$ convex and $f(y) \geq f(x) + \nabla f(x)^\top (y - x)$ for all x, y. (Tangent is global underestimator).
- **Second-Order Cond:** f twice diff. Convex iff $\nabla^2 f(x) \succeq 0 \, \forall x$.
- **Epigraph:** f is convex iff $\mathbf{epi} f = \{(x, t) \mid f(x) \leq t\}$ convex.
- **Sublevel Sets:** If f is convex, $C_\alpha = \{x \mid f(x) \leq \alpha\}$ is convex. (Converse NOT true).

Operations Preserving Convexity

- **Non-negative weighted sum:** $\alpha f + \beta g$ $(\alpha, \beta \geq 0)$.
- **Composition with Affine:** $f(Ax + b)$.
- **Pointwise Maximum:** $f(x) = \max\{f_1(x), \ldots, f_m(x)\}$.
- **Pointwise Supremum:** $f(x) = \sup_{y \in \mathcal{A}} f(x, y)$ (if $f(x, y)$ convex in x for each y). *Ex: Support function, Max eigenvalue.*
- **Minimization:** $g(x) = \inf_{y \in C} f(x, y)$ is convex if f is jointly convex in (x, y) and C is a convex set.
- **Perspective:** $g(x, t) = tf(x/t)$ with domain $\{(x, t) \mid x/t \in \mathbf{dom} f, t > 0\}$.
- **Scalar Composition:** $f(x) = h(g(x))$. Convex if (g convex, h convex & non-decreasing) or (g concave, h convex & non-increasing).
- **Vector Composition:** $f(x) = h(g_1(x), \ldots, g_k(x))$. Convex if: g_i convex, h convex & non-decreasing in each arg.

Quasiconvex Functions

- Definition: Domain convex and all sublevel sets $S_\alpha = \{x \mid f(x) \leq \alpha\}$ are convex.
- **Modified Jensen:** $f(\theta x + (1 - \theta)y) \leq \max(f(x), f(y))$.
- **First Order:** $f(y) \leq f(x) \implies \nabla f(x)^\top (y - x) \leq 0$.
- Examples: Linear-fractional functions, $\sqrt{|x|}$, Ratio of norms.
- *Note:* Sum of quasiconvex functions is NOT necessarily quasiconvex.

Log-Concave/Convex

- Log-concave: $f(x) > 0$ and $\log f(x)$ is concave. (e.g., Gaussian PDF, Indicator of convex set).
- Properties: Product of log-concave is log-concave. Integration (marginalization) preserves log-concavity. Convolution preserves log-concavity.

3. Matrix Calculus Identities

Vector Derivatives: $\nabla_x(x^\top Ax) = (A + A^\top)x$, $\nabla_x ||Ax - b||_2^2 = 2A^\top (Ax - b)$, $\nabla_x \|Ax\|_2 = \frac{A^\top Ax}{\|Ax\|_2}$.

Matrix Derivatives

- $\nabla_X \mathbf{Tr}(AX) = A^\top$ and $\nabla_X \mathbf{Tr}(X^\top AX) = (A + A^\top)X$.
- $\nabla_X \log\det X = X^{-1}$ (for $X \succ 0$) and $\nabla_X \det X = (\det X)X^{-\top}$.
- **Differential of Inverse:** $d(X^{-1}) = -X^{-1}(dX)X^{-1}$.
- **Hessian of Log-Det:** To find $\nabla^2 f(X)[V, V]$, consider $g(t) = \log\det(X + tV)$. $g'(t) = \mathbf{Tr}((X + tV)^{-1}V)$. $g''(t) = -\mathbf{Tr}((X + tV)^{-1}V(X + tV)^{-1}V)$. At $t = 0$: $-\mathbf{Tr}(X^{-1}VX^{-1}V)$. (Quadratic form is negative $\implies$ Concave).

Schur Complement

Let $X = \begin{bmatrix} A & B \\ B^\top & C \end{bmatrix}$ be a symmetric matrix. If $A \succ 0$, then $X \succeq 0$ if and only if the Schur complement $S = C - B^\top A^{-1}B \succeq 0$.

- **Application:** Converting nonlinear convex constraints into Linear Matrix Inequalities (LMIs).
- *Example:* $x^\top x \leq t$. Rewrite as $t - x^\top I^{-1}x \geq 0$. By Schur complement (with $A = I, B = x, C = t$), equiv to: $\begin{bmatrix} I & x \\ x^\top & t \end{bmatrix} \succeq 0$.

Matrix Fractional Function

$f(x, Y) = x^\top Y^{-1}x$ is jointly convex in (x, Y) for $Y \succ 0$. **Proof:** Epigraph is $\{(x, Y, t) \mid x^\top Y^{-1}x \leq t, Y \succ 0\}$. Using Schur complement, this is equivalent to the LMI: $\begin{bmatrix} Y & x \\ x^\top & t \end{bmatrix} \succeq 0$. Since the PSD cone is convex, the epigraph is convex.

4. Conjugate Functions

Definition & Properties

The conjugate of f is $f^*(y) = \sup_{x \in \mathbf{dom} f}(y^\top x - f(x))$.

- f^* is always convex (pointwise supremum of affine functions).
- **Fenchel's Inequality:** $f(x) + f^*(y) \geq x^\top y$.
- **Biconjugate:** $f^{**} = f$ if f is closed and convex.

Library of Conjugates

- **Affine:** $f(x) = ax + b \to f^*(y) = -b$ if $y = a$, ∞ otherwise.
- **Negative Log:** $f(x) = -\log x$ $(\mathbb{R}_{++}) \to f^*(y) = -1 - \log(-y)$ $(y < 0)$.
- **Exponential:** $f(x) = e^x \to f^*(y) = y\log y - y$ $(y \geq 0)$.
- **Entropy:** $f(x) = x\log x$ $(\mathbb{R}_+) \to f^*(y) = e^{y-1}$.
- **Inverse:** $f(x) = 1/x$ $(\mathbb{R}_{++}) \to f^*(y) = -2\sqrt{-y}$ $(y \leq 0)$.
- **Quadratic:** $f(x) = \frac{1}{2}x^\top Qx$ $(Q \succ 0) \to f^*(y) = \frac{1}{2}y^\top Q^{-1}y$.
- **Norm:** $f(x) = ||x|| \to f^*(y) = \iota_{B_*}(y)$ (Indicator of dual unit ball). E.g., $L_1 \to L_\infty$ ball; $L_2 \to L_2$ ball.
- **Norm Squared:** $f(x) = \frac{1}{2}||x||^2 \to f^*(y) = \frac{1}{2}||y||_*^2$.
- **Indicator:** $f(x) = \iota_C(x) \to f^*(y) = \sigma_C(y)$ (Support Function).
- **Log-Sum-Exp:** $f(x) = \log(\sum e^{x_i}) \to f^*(y) = \sum y_i \log y_i$ if $y \geq 0, \mathbf{1}^\top y = 1$ (Neg entropy on simplex).
- **Log-Determinant:** $f(X) = -\log\det X \to f^*(Y) = -n - \log\det(-Y)$ $(Y \prec 0)$.
- **Scaling/Composition:** $g(x) = \alpha f(x) \to g^*(y) = \alpha f^*(y/\alpha)$. $g(x) = f(Ax) \to g^*(y) = f^*(A^{-\top}y)$ (if A invertible). $g(x) = f(x + b) \to g^*(y) = f^*(y) - b^\top y$.

Support Function $\sigma_C(y)$: Defined as $\sigma_C(y) = \sup_{x \in C} y^\top x$.

5. Optimization Problems & Dual Derivations

1. Linear Programming (LP)

Primal: $\min c^\top x$ s.t. $Ax = b, x \geq 0$. *Derivation:* $L(x, \nu, \lambda) = c^\top x + \nu^\top (Ax - b) - \lambda^\top x = (c + A^\top \nu - \lambda)^\top x - b^\top \nu$. Minimizing over x requires coeff to be 0. **Dual:** $\max -b^\top \nu$ s.t. $A^\top \nu + c = \lambda \geq 0$.

2. Quadratic Programming (QP)

Primal: $\min \frac{1}{2}x^\top Px + q^\top x$ s.t. $Ax \leq b$ $(P \succ 0)$. *Derivation:* $L = \frac{1}{2}x^\top Px + q^\top x + \lambda^\top (Ax - b)$. $\nabla_x L = Px + q + A^\top \lambda = 0 \implies x^* = -P^{-1}(q + A^\top \lambda)$. Sub x^* back into L. **Dual:** $\max -\frac{1}{2}\lambda^\top (AP^{-1}A^\top)\lambda - (b + AP^{-1}q)^\top \lambda - \frac{1}{2}q^\top P^{-1}q$ s.t. $\lambda \geq 0$.

3. Second-Order Cone Programming (SOCP)

Primal: $\min f^\top x$ s.t. $\|A_i x + b_i\|_2 \leq c_i^\top x + d_i$, $i = 1, \ldots, m$. We introduce dual variables for each constraint i. Since the constraint represents membership in the second-order cone $\mathcal{K}_i = \{(y, t) \mid \|y\|_2 \leq t\}$, we associate dual variables (u_i, t_i) where u_i corresponds to the vector part $A_i x + b_i$ and t_i corresponds to the scalar part $c_i^\top x + d_i$. The dual variables must satisfy the dual cone condition. Since the second-order cone is self-dual, the condition is $\|u_i\|_2 \leq t_i$ for all i. The Lagrangian is formed by subtracting the inner product of the constraints and the dual variables from the objective $\mathcal{L}(x, u, t) = f^\top x - \sum_{i=1}^m \left(u_i^\top (A_i x + b_i) + t_i(c_i^\top x + d_i)\right)$. To minimize $\mathcal{L}(x, u, t)$ with respect to x, the coefficient of x must be 0. This yields the equality constraint $f = \sum_{i=1}^m (A_i^\top u_i + c_i t_i)$. Substituting into the Lagrangian leaves the constant term. **Dual SOCP:** $\max -\sum_{i=1}^m (b_i^\top u_i + d_i t_i)$ s.t. $\sum_{i=1}^m (A_i^\top u_i + c_i t_i) = f$ and $\|u_i\|_2 \leq t_i$ for all i.

4. Semidefinite Programming (SDP)

Primal: $\min \mathrm{tr}(CX)$ s.t. $\mathrm{tr}(A_i X) = b_i$, $i = 1, \ldots, m$, $X \succeq 0$. We introduce scalar dual variables y_i for the linear equalities and a matrix dual variable $S \succeq 0$ for the positive semidefinite cone constraint. The Lagrangian is formed by adding the equality constraints and subtracting the conic inner product from the objective: $\mathcal{L}(X, y, S) = \mathrm{tr}(CX) + \sum_{i=1}^m y_i(b_i - \mathrm{tr}(A_i X)) - \mathrm{tr}(SX)$. Grouping terms involving X gives $\mathcal{L}(X, y, S) = b^\top y + \mathrm{tr}\left((C - \sum_{i=1}^m y_i A_i - S)X\right)$. To minimize $\mathcal{L}$ with respect to X, the coefficient matrix of X must vanish (otherwise the minimum is $-\infty$). This yields the equality $C - \sum_{i=1}^m y_i A_i - S = 0$, or equivalently $S = C - \sum_{i=1}^m y_i A_i$. Substituting this back leaves the constant term $b^\top y$. We maximize this subject to the dual cone constraint $S \succeq 0$. **Dual:** $\max b^\top y$ s.t. $\sum_{i=1}^m y_i A_i \preceq C$.

5. Geometric Programming (GP)

Primal: $\min \log \sum_{k \in K_0} e^{a_k^\top x + b_k}$ s.t. $\log \sum_{k \in K_i} e^{a_k^\top x + b_k} \leq 0$, $i = 1, \ldots, m$. We introduce a dual variable vector $\nu \geq 0$, where each component ν_k corresponds to an exponential term $e^{a_k^\top x + b_k}$. The Lagrangian is derived using the conjugate of the log-sum-exp function (which is the negative entropy). Minimizing the Lagrangian with respect to x requires the gradients to sum to zero, yielding the orthogonality condition $\sum_k \nu_k a_k = 0$. The objective term requires a normalization condition $\sum_{k \in K_0} \nu_k = 1$. Substituting these conditions back, the dual function maximizes the linear term $b^\top \nu$ minus the entropy of ν, adjusted for the block sums $\lambda_i = \sum_{k \in K_i} \nu_k$. **Dual:** $\max b^\top \nu - \sum_k \nu_k \log(\nu_k/\lambda_{i(k)})$ s.t. $\sum_k \nu_k a_k = 0, \sum_{k \in K_0} \nu_k = 1, \nu \geq 0$.

6. Duality Theory

Lagrangian
$L(x,\lambda,\nu) = f_0(x) + \sum \lambda_i f_i(x) + \sum \nu_i h_i(x)$. Domain $\mathcal{D} = \mathbf{dom} f_0 \cap (\cap \mathbf{dom} f_i) \cap (\cap \mathbf{dom} h_i)$. λ_i (inequality mult) must be ≥ 0.
Dual Function
$g(\lambda,\nu) = \inf_{x \in \mathcal{D}} L(x,\lambda,\nu)$. **Properties:** g is always concave (even if primal is nonconvex), as it is the pointwise infimum of affine functions of (λ,ν). **Lower Bound:** For any $\lambda \geq 0$, $g(\lambda,\nu) \leq p^*$.
The Dual Problem
$\max g(\lambda,\nu) \quad \text{s.t.} \quad \lambda \geq 0$. This is always a convex optimization problem. Optimal value denoted d^*.
Duality Gap
Weak Duality: $d^* \leq p^*$ always. **Strong Duality:** $d^* = p^*$. Usually holds for convex problems under constraint qualifications.
Slater's Condition (Constraint Qualification)
Strong duality holds if the problem is **Convex** AND **Strictly Feasible**. Strict Feasibility: $\exists x \in \mathbf{relint}(\mathcal{D})$ s.t. $f_i(x) < 0$ (for non-affine constraints) and $Ax = b$. *Note:* Linear inequalities $Ax \leq b$ do NOT require strict inequality for Slater's. Feasibility is enough. *For SDPs:* Strong duality holds if Primal strictly feasible ($X \succ 0$) OR Dual strictly feasible ($S \succ 0$).
KKT Conditions
For a problem with differentiable functions. 1. **Primal Feasibility:** $f_i(x) \leq 0, h_i(x) = 0$. 2. **Dual Feasibility:** $\lambda \geq 0$. 3. **Complementary Slackness:** $\lambda_i f_i(x) = 0$ (Either constraint active or multiplier zero). 4. **Stationarity:** $\nabla f_0(x) + \sum \lambda_i \nabla f_i(x) + \sum \nu_i \nabla h_i(x) = 0$.
Theorem:
- If strong duality holds, any pair of primal/dual optimal points satisfies KKT.
- If problem is convex, any points satisfying KKT are primal/dual optimal.

Sensitivity Analysis
Consider perturbed problem with limits u_i, v_i: $p^*(u,v) = \min f_0(x) \quad \text{s.t.} \quad f_i(x) \leq u_i, h_i(x) = v_i$. **Global:** $p^*(u,v) \geq p^*(0,0) - \lambda^{*\top} u - \nu^{*\top} v$. **Local:** If differentiable, $\lambda_i^* = -\frac{\partial p^*(0,0)}{\partial u_i}$ and $\nu_i^* = -\frac{\partial p^*(0,0)}{\partial v_i}$. Interpretation: λ_i^* is the "shadow price" – how much objective improves if constraint relaxed.

7. Classic Reformulations & Tricks

A. Robust LP → SOCP
Problem: $\min c^\top x \quad \text{s.t.} \quad a_i^\top x \leq b_i$ for all $a_i \in \mathcal{E}_i$. Uncertainty set (Ellipsoid): $\mathcal{E}_i = \{\bar{a}_i + P_i u \mid ||u||_2 \leq 1\}$. Constraint is: $\sup_{a_i \in \mathcal{E}_i} a_i^\top x \leq b_i$. Derivation: $\sup_{||u|| \leq 1} (\bar{a}_i + P_i u)^\top x = \bar{a}_i^\top x + \sup_{||u|| \leq 1} (P_i^\top x)^\top u$. Recall definition of dual norm: $\sup_{||u|| \leq 1} z^\top u = ||z||_*$. Dual of L_2 is L_2. So, $\sup(\dots) = ||P_i^\top x||_2$. Resulting Constraint: $\bar{a}_i^\top x + ||P_i^\top x||_2 \leq b_i$. This is exactly an SOC constraint.
B. Stochastic LP → SOCP
Constraint $a_i^\top x \leq b_i$ must hold with probability $\eta \geq 0.5$. Assume $a_i \sim \mathcal{N}(\bar{a}_i, \Sigma_i)$. Then $u = a_i^\top x$ is Gaussian with mean $\bar{u} = \bar{a}_i^\top x$ and var $\sigma^2 = x^\top \Sigma_i x$. Condition: $P(u \leq b_i) \geq \eta \iff \Phi(\frac{b_i - \bar{u}}{\sigma}) \geq \eta$. $\iff b_i - \bar{a}_i^\top x \geq \Phi^{-1}(\eta) ||\Sigma_i^{1/2} x||_2$. Result: $\bar{a}_i^\top x + \Phi^{-1}(\eta) ||\Sigma_i^{1/2} x||_2 \leq b_i$. (SOCP).
C. Minimizing Spectral Radius
Problem: $\min_x \rho(A(x))$ where $A(x) \in S^n$ depends affinely on x. Recall $\rho(A) = \max_i |\lambda_i(A)| = \max(\lambda_{\max}(A), \lambda_{\max}(-A))$. Constraint $\rho(A) \leq t \iff -t \leq \lambda_i \leq t \iff -tI \preceq A \preceq tI$. SDP Formulation:

$$\min t \quad \text{s.t.} \quad A(x) \preceq tI, \quad -A(x) \preceq tI$$

D. Matrix Norm Minimization
Problem: $\min_x ||A(x)||_2$ (Spectral norm = max singular value). Constraint $||A||_2 \leq t \iff A^\top A \preceq t^2 I$. Using Schur complement on $\begin{bmatrix} tI & A \\ A^\top & tI \end{bmatrix} \succeq 0$: $tI - A^\top (tI)^{-1} A \succeq 0 \iff t^2 I - A^\top A \succeq 0$. SDP Formulation: $\min t \quad \text{s.t.} \quad \begin{bmatrix} tI & A(x) \\ A(x)^\top & tI \end{bmatrix} \succeq 0$.

E. Chebyshev Center
Find center x_c and radius r of largest ball inside Polyhedron $P = \{x \mid a_i^\top x \leq b_i\}$. Condition: $x \in B(x_c, r) \implies a_i^\top x \leq b_i$. $\sup_{||u|| \leq 1} a_i^\top (x_c + ru) \leq b_i \iff a_i^\top x_c + r||a_i||_2 \leq b_i$. LP Formulation: $\max r \quad \text{s.t.} \quad a_i^\top x_c + r||a_i||_2 \leq b_i$.
F. Analytic Center
For a set of inequalities $a_i^\top x \leq b_i$. The analytic center is the solution to $\min - \sum \log(b_i - a_i^\top x)$. This is an unconstrained convex problem (domain is interior of polyhedron). Newton's method works very well here.

8. Solved Exercise Library

Ex 1: LASSO Dual (HW3)
Primal: $\min_w \frac{1}{2}||Xw - y||_2^2 + \lambda ||w||_1$. *Recipe Step 1: Reformulate.* Introduce $z = Xw - y$. Primal: $\min \frac{1}{2}||z||_2^2 + \lambda ||w||_1 \quad \text{s.t.} \quad z = Xw - y$. *Step 2: Lagrangian.* Multiplier ν for equality. $L(w,z,\nu) = \frac{1}{2}||z||^2 + \lambda ||w||_1 + \nu^\top (Xw - y - z)$. *Step 3: Minimize L.* Separable in z and w. Min over z: $\nabla_z L = z - \nu = 0 \implies z = \nu$. Term becomes: $\frac{1}{2}||\nu||^2 - ||\nu||^2 - \nu^\top y = -\frac{1}{2}||\nu||^2 - \nu^\top y$. Min over w: $\inf_w(\lambda ||w||_1 + (X^\top \nu)^\top w)$. This relates to conjugate of norm. $\inf_x(c^\top x + ||x||)$ is bounded only if dual norm of cost is small. Analytically: $\sum(\lambda |w_i| + (X^\top \nu)_i w_i)$. If $|(X^\top \nu)_i| > \lambda$, we can send $w_i \to -\infty$ to get $-\infty$. Condition: $||X^\top \nu||_\infty \leq \lambda$. If holds, min is 0. *Step 4: Dual.* $\max_\nu -\frac{1}{2}||\nu||_2^2 - \nu^\top y \quad \text{s.t.} \quad ||X^\top \nu||_\infty \leq \lambda$ This is a Box-Constrained QP.
Ex 2: SVM / Hinge Loss Dual (HW2)
Primal: $\min_{w,\xi} \frac{1}{n} \sum \xi_i + \frac{\tau}{2}||w||^2 \quad \text{s.t.} \quad \xi_i \geq 1 - y_i w^\top x_i, \xi_i \geq 0$. *Lagrangian:* multipliers $\alpha_i \geq 0$ (margin), $\beta_i \geq 0$ (non-neg). $L = \sum \frac{1}{n}\xi_i + \frac{\tau}{2}||w||^2 - \sum \alpha_i(\xi_i - 1 + y_i w^\top x_i) - \sum \beta_i \xi_i$. *Stationarity w:* $\nabla_w L = \tau w - \sum \alpha_i y_i x_i = 0 \implies w = \frac{1}{\tau} \sum \alpha_i y_i x_i$. *Stationarity ξ:* $\nabla_{\xi_i} L = \frac{1}{n} - \alpha_i - \beta_i = 0 \implies \alpha_i + \beta_i = 1/n$. Since $\beta_i \geq 0$, this implies $0 \leq \alpha_i \leq 1/n$. *Substitute:* Plug w back into L. Term 1: $\frac{\tau}{2}||w||^2 = \frac{1}{2\tau}||\sum \alpha_i y_i x_i||^2$. Term 2: $-\sum \alpha_i y_i w^\top x_i = -w^\top(\tau w) = -\tau ||w||^2 = -\frac{1}{\tau}||\sum \alpha_i y_i x_i||^2$. Sum is $-\frac{1}{2\tau}\sum_{i,j} \alpha_i \alpha_j y_i y_j x_i^\top x_j$. Linear term: $\sum \alpha_i$. **Dual:** $\max_\alpha \sum \alpha_i - \frac{1}{2\tau}\sum_{i,j} \alpha_i \alpha_j y_i y_j x_i^\top x_j \quad \text{s.t.} \quad 0 \leq \alpha_i \leq \frac{1}{n}$.
Ex 3: Water-filling (Lecture 3)
Problem: $\min -\sum \log(x_i + \alpha_i) \quad \text{s.t.} \quad \sum x_i = 1, x \geq 0$. KKT Conditions: 1. Stationarity: $-\frac{1}{x_i + \alpha_i} + \nu - \lambda_i = 0$. 2. Slackness: $\lambda_i x_i = 0, \lambda_i \geq 0$. Solve for x_i: $x_i + \alpha_i = 1/(\nu - \lambda_i)$. If $\nu < 1/\alpha_i$, then $1/(\nu - \lambda_i)$ must equal α_i (if $x = 0$) or higher. Essentially: $x_i = 1/\nu - \alpha_i$. If this value is negative, constraints force $x_i = 0$. Result: $x_i^* = \max(0, 1/\nu - \alpha_i)$. Find ν by solving $\sum \max(0, 1/\nu - \alpha_i) = 1$. This is monotonic piecewise linear.
Ex 4: Boolean Relaxation (HW2)
Problem: $\min c^\top x \quad \text{s.t.} \quad Ax \leq b, x_i \in \{0,1\}$. Hard constraint: $x_i \in \{0,1\} \iff x_i(1 - x_i) = 0$. Lagrangian Relaxation: Relax explicit constraints $Ax \leq b$ or equality constraints. Consider $L(x,\nu) = c^\top x + \sum \nu_i(x_i^2 - x_i)$. (Using quadratic equality). Dual function $g(\nu) = \inf_x L(x,\nu) = \inf_x \sum(\nu_i x_i^2 + (c_i - \nu_i)x_i)$. This is separable. For each i, we minimize a parabola. If $\nu_i > 0$, parabola convex $\to x_i^* = -(c_i - \nu_i)/(2\nu_i)$. Value is finite. If $\nu_i < 0$, parabola concave $\to$ unbounded ($-\infty$). Dual Problem: Maximize the sum of these minima subject to $\nu \geq 0$. This gives a lower bound on the Boolean LP.
Ex 5: Analytic Center of LMI
Problem: $\min \phi(x) = -\log\det(B - \sum_{i=1}^n x_i A_i)$. Domain: $F(x) = B - \sum x_i A_i \succ 0$. **Gradient:** Chain rule with $\nabla \log\det X = X^{-1}$. $d\phi = -\mathbf{Tr}(F(x)^{-1} dF(x)) = -\mathbf{Tr}(F^{-1}(-\sum dx_i A_i)) = \sum dx_i \mathbf{Tr}(F^{-1} A_i)$. So $\nabla \phi_i = \mathbf{Tr}(F(x)^{-1} A_i)$. **Hessian:** Differential of X^{-1} is $-X^{-1} dX X^{-1}$. $d(\nabla \phi_i) = \mathbf{Tr}(d(F^{-1}) A_i) = \mathbf{Tr}(-F^{-1} dF F^{-1} A_i) = \mathbf{Tr}(F^{-1}(\sum_j dx_j A_j) F^{-1} A_i) = \sum_j dx_j \mathbf{Tr}(F^{-1} A_j F^{-1} A_i)$. So $\nabla^2 \phi_{ij} = \mathbf{Tr}(F(x)^{-1} A_i F(x)^{-1} A_j)$. This matrix is always PSD (Gram matrix structure).
Ex 6: Regularized Least Squares (HW2)
Problem: $\min_x ||Ax - b||_2^2 + ||x||_1$. Dual Derivation: Same variables as LASSO. Primal can be seen as $\min ||y||_2^2 + ||x||_1 \quad \text{s.t.} \quad y = Ax - b$. $L = ||y||_2^2 + ||x||_1 + \nu^\top(Ax - b - y)$. $\inf_y(||y||_2^2 - \nu^\top y)$. Conj of square is square/4. Value: $-\frac{1}{4}||\nu||^2$. $\inf_x(||x||_1 + (A^\top \nu)^\top x) \to$ Constraint $||A^\top \nu||_\infty \leq 1$. Dual: $\max -b^\top \nu - \frac{1}{4}||\nu||^2 \quad \text{s.t.} \quad ||A^\top \nu||_\infty \leq 1$.
Ex 7: Max Entropy
Problem: $\min \sum x_i \log x_i \quad \text{s.t.} \quad Ax \leq b, \mathbf{1}^\top x = 1$. Conjugate of entropy $f(x) = x \log x$ is $f^*(y) = e^{y-1}$. Lagrangian: $L = \sum x_i \log x_i + \lambda^\top(Ax - b) + \nu(\mathbf{1}^\top x - 1)$. Stationarity: $1 + \log x_i + A_i^\top \lambda + \nu = 0 \implies x_i = e^{-(A^\top \lambda)_i - \nu - 1}$. Dual Function: Subst x_i. $\sum x_i \log x_i = \sum x_i(-(A^\top \lambda)_i - \nu - 1)$. $L = -b^\top \lambda - \nu + \sum x_i(\dots + (A^\top \lambda)_i + \nu) = -b^\top \lambda - \nu - \sum x_i$. Note $\sum x_i = \sum e^{-(A^\top \lambda)_i} e^{-\nu - 1}$. Maximize over ν analytically. Result connects to Geometric Programming duals (Log-Sum-Exp).

9. Algorithms

Newton's Method
For minimizing convex $f(x)$. **Search Direction:** $\Delta x_{nt} = -\nabla^2 f(x)^{-1} \nabla f(x)$. Affine Invariant. Solves quadratic approximation exactly. **Newton Decrement:** $\lambda(x) = (\nabla f^\top \nabla^2 f^{-1} \nabla f)^{1/2}$. Stopping Criterion: $\lambda^2/2 \leq \epsilon$. **Convergence:** Phase 1 (Damped): Linear convergence. Phase 2 (Pure): Quadratic convergence (doubles digits). Steps to precision ϵ: $c + \log\log(1/\epsilon)$.
Barrier Method (Interior Point)
Problem: $\min f_0(x) \quad \text{s.t.} \quad f_i(x) \leq 0$. **Log Barrier:** $\phi(x) = -\sum_{i=1}^m \log(-f_i(x))$. Parameter $t > 0$. **Central Path:** $x^*(t) = \operatorname{argmin} t f_0(x) + \phi(x)$. Optimality condition for $x^*(t)$: $\nabla f_0 + \sum \frac{1}{-t f_i} \nabla f_i = 0$. Identify dual points $\lambda_i^*(t) = -1/(t f_i(x))$. **Duality Gap:** Gap is exactly m/t. **Algorithm:** 1. Given strict feasible x, $t := t^{(0)}$, $\mu > 1$. 2. **Centering Step:** Compute $x^*(t)$ using Newton's method starting at x. 3. **Update:** $x := x^*(t)$. 4. **Stopping:** If $m/t < \epsilon$, return x. 5. **Increase:** $t := \mu t$. Total Newton steps scales with $\sqrt{m}$.
Backtracking Line Search
Used to find step size s for descent direction Δx. Parameters: $\alpha \in (0, 0.5), \beta \in (0,1)$. 1. Set $s = 1$. 2. While $f(x + s\Delta x) > f(x) + \alpha s \nabla f(x)^\top \Delta x$: $s := \beta s$. 3. Update $x := x + s\Delta x$.
Quasiconvex Bisection
Problem: $\min f_0(x) \quad \text{s.t.} \quad f_i(x) \leq 0$ (f_0 quasiconvex). Algorithm: 1. Interval $[l,u]$ containing p^*. 2. Solve convex feasibility: $\phi_t(x) \leq 0, f_i(x) \leq 0$ where ϕ_t is sublevel set at $t = (l+u)/2$. 3. If feasible, $u := t$. Else $l := t$. 4. Stop when $u - l \leq \epsilon$. Iterations: $\lceil \log_2((u_{init} - l_{init})/\epsilon) \rceil$.